# Zusammenhang zwischen perzeptiven Dimensionen und Störungsursachen bei super-breitbandiger Sprachübertragung

Sebastian Möller[1,2], Tobias Hübschen[3], Gabriel Mittag[1], Gerhard Schmidt[3]

[1] *Quality and Usability Lab, TU Berlin, E-Mail: sebastian.moeller@tu-berlin.de; gabriel.mittag@tu-berlin.de*

[2] *Speech and Language Technology, DFKI Berlin*

[3] *DSS, Christian-Albrechts-Universität zu Kiel, E-Mail: thu@tf.uni-kiel.de; gus@tf.uni-kiel.de*

## Einleitung

Die Bestimmung der Qualität übertragener Sprache ist seit jeher ein wichtiger Bestandteil der Planung, Implementierung und Überwachung von Sprachkommunikationsdiensten. Mit der Einführung IP-basierter Übertragung ergibt sich jedoch ein Paradigmenwechsel, da das übertragbare Frequenzband nicht mehr an das klassische Telefonieband (300-3400 Hz, Schmalband/NB) gebunden ist. Vielmehr wird durch die Einführung neuer Codierer auch breitbandige (50-7000 Hz/WB), super-breitbandige (20-14000 Hz/SWB) oder vollbandige (0-20000 Hz/FB) Übertragung möglich. Dadurch ändert sich der Charakter der übertragenen Sprache, und somit das Sprach-Hörereignis, im Allgemeinen in eine positive Richtung. Im Gegenzug können aber technische Ursachen des Übertragungskanals, wie niederbitratige Codierer, Paketverluste, Rauschen etc., die wahrgenommene Qualität wieder reduzieren.

Zur Diagnose der Qualität übertragener Sprache bieten sich analytische Hörversuche an, bei denen Probanden verschiedene perzeptive Dimensionen des Gehörten bewerten. Hierzu wurde von Wältermann [1][2] zunächst ein 3-dimensionales und anschließend ein 4-dimensionales Verfahren vorgeschlagen, welches die Dimensionen „Klangverfärbung“, „Rauschhaftigkeit“, „Diskontinuierlichkeit“ und (in der 4-dimensionalen Version) „nicht-optimale Lautheit“ von Versuchspersonen auf 5-stufigen Skalen bewerten lässt. Unklar ist bislang, inwieweit sich durch diese Bewertungen der perzeptiven Dimensionen auch die technischen Störungsursachen ermitteln lassen.

Zur Klärung dieser Frage wurde das Wältermannsche Verfahren auf 8 Datenbanken angewendet, die super-breitbandige Sprache mit verschiedenen, teilweise kontrolliert eingestellten, Störungen enthalten. Die Bewertungen der Probanden wurden dahingehend klassifiziert, ob eine Dimension auf mehr als die Hälfte des Maximalwertes abfällt, und daraus wurden Klassen von Störungsursachen für jede perzeptive Dimension erstellt. In den kommenden Abschnitten wird zunächst ein Überblick über die Datenbanken gegeben, anschließend werden die Analysen präsentiert, und abschließend die Ergebnisse in tabellarischer Form zusammengefasst.

## Datenbanken

Die 8 Datenbanken bestehen aus gestörten Sprachaufnahmen von 4-8 s Länge, einer textlichen Beschreibung der Störungsarten, sowie den gemittelten Urteilen der Probanden. Zwei der Datenbanken wurden noch mit dem 3-dimensionalen Verfahren (also ohne die Dimension “nicht-optimale Lautheit“) bewertet [2], die übrigen mit dem 4-dimensionalen Verfahren [3]. Das Abhören der Sprachproben erfolgte in ruhiger Umgebung über Kopfhörer durch muttersprachliche Hörer. Die Datenbanken wurden freundlicherweise von der Deutsche Telekom AG (DTAG, Deutsch), der Fa. Orange (Französisch) und der Fa. SwissQual (Schweizerdeutsch) zur Verfügung gestellt; die TUBDIS-Datenbank (Deutsch) stammte noch aus den Versuchen von Wältermann an der TU Berlin. Wichtige Eigenschaften der Datenbanken sind in Tab. 1 zusammengefasst.

**Tabelle 1:** Übersicht der verwendeten Datenbanken. Angegeben sind die Anzahl der Sprecher (männlich, weiblich) der Sprachproben, die Anzahl der Bewertungen pro Störungsart, die in den Datenbanken enthaltenen Frequenzbänder, sowie die Anzahl und Art der verwendeten Störungen. Bezeichner der Störungsarten: C: Codecs, T: Codec-Tandems, L: Leitungsrauschen, H: Hintergrundgeräusch sendeseitig, M: signalkorreliertes Rauschen erzeugt mittels einer MNRU [4], P: Paketverluste, Cl: Temporal Clipping, O: Overload, F: Bandpass-Filter, V: Verstärkungsänderungen, RE/RA: elektrische bzw. akustische Aufzeichnungen aus realen Netzen, W: Time Warping, A: Automatic Gain Conrol

| **Datenbank** | **Eigenschaften** | | | |
|---|---|---|---|---|
| | Anz. Sprech. | Bewert. /Stör. | Band-breite | Störungsanz. und Art |
| **DTAG 1** | 1m, 1w | 40 | NB, WB | 66: C, T, L, P, F |
| **DTAG 2** | 1m, 1w | 48 | NB, WB | 76: H, C, P, F, M |
| **DTAG 3** | 2m, 2w | 80 | NB-SWB | 54: C, RE, RA, H, M, F, V, Cl, W, L, A |
| **Orange 1** | 2m, 2w | 18 | NB-SWB | 56: H, L, M, RA, F, V, Cl, C, P, W |
| **SwissQual 1** | 2m, 2w | 96 | NB-SWB | 30: C, H, F, M, N, Cl, O, V, RA, RE |
| **SwissQual501** | 2m, 2w | 48 | NB-SWB | 50: C, T, H, V, F, M, P, Cl, RA, RE |
| **SwissQual502** | 2m 2w | 48 | NB-SWB | 50: C, T, H, V, F, M, P, Cl, RA, RE |
| **TUBDIS** | 1m, 1w | 35-41 | SWB | 20: L, V, C, Cl, F, P |

## Analysen

Für jede der in Tab. 1 gelisteten Datenbanken sollen nun technische Störungsursachen identifiziert werden, welche zu einer Beeinträchtigung der Bewertung in einer oder mehreren

perzeptiven Dimensionen geführt haben könnten. Hierzu stehen die Beschreibungen der Störungsarten zur Verfügung, wie sie in den Datenbanken vorliegen. Die Beschreibung ist insbesondere hilfreich, wenn es sich um simulierte Störungen handelt, da bei diesen die Verarbeitungsschritte mehr oder weniger genau dokumentiert sind. Im Gegenzug dazu ist es bei den Sprachdateien, welche in realen Netzen (elektrisch) und/oder über reale Endgeräte (akustisch) aufgezeichnet wurden, normalerweise nicht möglich, eine technische Störungsursache zu identifizieren; diese Störungen sind in Tab. 1 mit RA und RE gekennzeichnet.

Zur Bestimmung einer perzeptiven „Beeinträchtigung" muss zunächst ein Kriterium gewählt werden, welches für alle unterschiedlichen Datenbanken anwendbar ist. Da die Bewertungen der perzeptiven Dimensionen auf Skalen mit 5 Attributen erfolgte, und die Bewertungen zu arithmetischen Mittelwerten (5=optimal, 1=schlecht) über alle Sprachdateien einer Störungsklasse aggregiert wurden, wurde als Kriterium eine mittlere Bewertung von ≤ 3,0 für die entsprechende perzeptive Dimension gewählt. Es sei erwähnt, dass nicht alle perzeptiven Dimensionen gleich stark auf die Gesamtqualität wirken, d.h. dass ggf. mit unterschiedlichen Schwellwerten pro Dimension gerechnet werden könnte. Da allerdings die Gewichtungen der perzeptiven Dimensionen bezüglich der Gesamtqualität hier zunächst unbekannt sind, und da auch nicht davon auszugehen ist, dass sie bei jeder Datenbank gleich sind, wird hier ein einziger Schwellwert für alle Dimensionen gewählt.

In Tab. 2 wurden nun alle Störungsarten aufgeführt, die bei einzelnen Datenbanken zu einer perzeptiven Beeinträchtigung einer Dimension geführt haben. Dabei wurde insofern verallgemeinert, dass nicht jede einzelne Störungsart genau aufgeführt wurde, sondern dass Klassen ähnlicher Störungsarten gebildet wurden, die im Allgemeinen zu einem mittleren Urteil ≤ 3,0 führten. In der letzten Spalte von Tab. 2 wurden die Störungsklassen aufgelistet, die in allen oder fast allen Datenbanken gleichermaßen identifiziert wurden; Störungsarten, die nur in einzelnen Fällen zu einer wahrgenommenen Beeinträchtigung dieser Dimension führten, sind in dieser Spalte kursiv gesetzt.

**Tabelle 2:** Übersicht der verwendeten Datenbanken. Bezeichner der Störungsarten: C: Codecs, T: Codec-Tandems, L: Leitungsrauschen, H: Hintergrundgeräusch sendeseitig, M: signalkorreliertes Rauschen erzeugt mittels einer MNRU [4], P: Paketverluste, Cl: Temporal Clipping, O: Overload, F: Bandpass-Filter, V: Verstärkungsänderungen, RE/RA: elektrische bzw. akustische Aufzeichnungen aus realen Netzen, W: Time Warping, A: Automatic Gain Control

| **Perzeptive Dimension** | **Datenbanken** | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| | DTAG 1 | DTAG 2 | DTAG 3 | Orange 1 | Swiss Qual 1 | Swiss Qual 501 | Swiss Qual 502 | TUBDIS | alle |
| **Klangverfärbung** | C(NB), T(NB), F | C(NB), F, M | F, C(NB), T(NB), M+H | M, F, RA, C(NB), O, Cl (>2%) | F, C(NB) | F, C(NB und WB), M | F, M, C(NB), RA | F, P | C(NB), T(NB), F, RA, *P*, *Cl*, *O* |
| **Rauschhaftigkeit** | H, C (G.726), T | H, M | H, M, L | M, H, L | H, M | H, M, RA | H, L, M, RA, | L, P | M, L, H, *C(G.726)*, *C(AMR-NB), P* |
| **Diskuntinuierlich-keit** | C, T, Cl | P, M | Cl | Cl (>10%) | Cl | Cl, P, RE | Cl, O | Cl, P | P, Cl, O, A, RE, *C(NB), T(NB), M* |
| **Nicht-optimale Lautheit** | -- | -- | V, T(triple), C | V, A | V | V | V | V | V, A |

Eine genaue Inspektion der Tab. 2 zeigt viele erwartbare Ergebnisse:

- *Klangverfärbung*: In einem experimentellen Kontext, in dem neben schmal- auch breitbandige oder super-breitbandige Sprachaufnahmen präsentiert werden, werden schmalbandige Aufnahmen normalerweise als klanglich verfärbt wahrgenommen. Insbesondere gilt dies in Kombination mit niederbitratigen Codecs oder Codec-Tandems. Auch Bandpass-Filterungen rufen Klangverfärbungen hervor, oder Frequenzverzerrungen, wie sie durch die akustisch-elektrischen Übertragungsfunktionen von Endgeräten entstehen. Nicht erwartet waren hingegen die in manchen Fällen auftretenden Klangverfärbungen bei Paketverlusten (P), bei *Temporal Clipping* (Cl) oder bei Übersteuerungen (O).
- *Rauschhaftigkeit*: Hier war wiederum zu erwarten, dass Leitungsrauschen, Hintergrundgeräusche sowie signalkorreliertes Rauschen, wie es eine *Modulated Noise Reference Unit* (MNRU, [4]) hervorruft, diese Dimension beeinträchtigen. Nicht erwartet war hingegen, dass auch der ADPCM-Kodierer nach ITU-T Rec. G.726 sowie der AMR-NB-Kodierer, beide insbesondere bei niedrigen Bitraten, Rauschhaftigkeit hervorrufen können. In wenigen Fällen führten auch Paketverluste zu wahrgenommener Rauschhaftigkeit, wahrscheinlich wegen der mit den Verlusten einhergehenden Signalverzerrungen.
- *Diskontinuierlichkeit*: Erwartungsgemäß wird diese durch Paketverluste, durch *Temporal Clipping*, durch Übersteuerungen, durch automatische Pegelregelungen (AGC) sowie durch Übertragungsfehler in realen Netzen hervorgerufen. Allerdings – und nicht erwartet – zeigte sich Diskontinuierlichkeit auch bei einigen niederbitratigen Codecs (bzw. Codec-Tandems), sowie manchmal auch bei signalkorreliertem Rauschen.

- *Nicht-optimale Lautheit*: Diese wird bei nicht-optimaler Pegeleinstellung sowie bei Pegelvariationen und bei AGC beobachtet.

Die Zusammenfassung der technischen Ursachen in der letzten Spalte von Tab. 2 zeigt, dass für jede perzeptive Beeinträchtigung einer Dimension unterschiedliche technische Störungsursachen in Frage kommen. Umgekehrt zeigt ein Vergleich der technischen Störungsursachen über die perzeptiven Dimensionen hinweg aber auch, dass die gleiche technische Störungsursache mehr als nur eine perzeptive Dimension beeinträchtigen kann. Beispielsweise können (insbesondere niederbitratige) Kodierer sowohl Klangverfärbungen als auch Diskontinuierlichkeit hervorrufen, in manchen Fällen sogar auch Rauschhaftigkeit. Paketverluste zeigen sich durch Diskontinuierlichkeit, aber manchmal auch durch Klangverfärbungen. Insbesondere das signalkorrelierte Rauschen, welches in den Experimenten (realitätsfern) mittels MNRU generiert wurde, zeigt sich häufig in der Beeinträchtigung aller vier perzeptiver Dimensionen.

Durch die binäre Klassifikation von „Beeinträchtigung“ (anhand des Schwellwertes von 3,0) werden weitere Informationen unterdrückt, die in der praktischen Anwendung der Diagnose von Bedeutung sein können. So kann über den genauen numerischen Beurteilungswert die Relevanz der Dimension auch für das Gesamtqualitätsurteil abgeschätzt werden, und damit entschieden werden, welche Beeinträchtigungen prioritär zu reduzieren sind. Umgekehrt kann Vorwissen über das technische Set-Up des Übertragungsnetzes helfen, aus den perzeptiven Dimensionen mögliche technische Ursachen einzugrenzen. Solches Vorwissen kann z.B. den verwendeten Kodierer betreffen, oder es könnte aus dem Jitter-Buffer Information über mögliche Paketverluste abgeleitet werden.

## Diskussion

Es zeigt sich, dass das auditive Verfahren sehr gut zur Klassifikation eingesetzt werden kann, auch wenn eine eindeutige technische Ursachenanalyse damit nicht möglich ist. Insbesondere gibt es Störungsursachen wie bspw. Codecs, die mehr als eine perzeptive Dimension betreffen. Allerdings lässt sich aus der perzeptiven Dimensionsanalyse auch für diese Ursachen eine Diagnose der perzeptiv dominanten Effekte erstellen.

Die Ergebnisse sind insbesondere auch für die instrumentelle Schätzung von perzeptiven Dimensionen und technischen Störungsursachen relevant. Solche Schätzer können intrusiv – auf Basis des Eingangs- und Ausgangssignals der Übertragungsstrecke (bspw. das DIAL-Modell von Côté [5], welches zu einem standardisierten Schätzer weiterentwickelt werden soll [6]) – oder auch nicht-intrusiv, d.h. allein auf Basis des gestörten Sprachsignals [7][8][9], einzelne perzeptive Dimensionen schätzen. Die hier vorgelegten Ergebnisse zeigen, welche technische Störungsursachen solche Schätzer erfassen müssen, um allgemeingültige Schätzwerte zu liefern.

## Literatur

[1] Wältermann, M., Raake, A., Möller, S.: Quality Dimensions of Narrowband and Wideband Speech Transmission, Acta Acustica united with Acustica 96(6), pp. 1090-1103, 2010

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

[3] ITU-T Contr. COM 12-C195: Draft Requirement Specification for P.AMD (Perceptual Approaches for Multidimensional Analysis), Source: Deutsche Telekom AG, Int. Telecomm. Union, CH-Geneva, Jan. 2011

[4] ITU-T Rec. P.810: Modulated Noise Reference Unit (MNRU), Int. Telecomm. Union, CH-Geneva, 1996

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

[6] ITU-T Contr. SG12-C.303: P.AMD Set A Updated Performance Results of the Noisiness Dimension, Source: Deutsche Telekom AG, Opticom GmbH, Int. Telecomm. Union, CH-Geneva, Nov. 2018

[7] ITU-T Contr. SG12-C.42: First Possible P.SAMD Indicators for the Estimation of Coloration, Source: Deutsche Telekom AG, Int. Telecomm. Union, CH-Geneva, 2017

[8] ITU-T Contr. SG12-C.43: First Possible P.SAMD Indicators for the Estimation of Noisiness, Source: Deutsche Telekom AG, Int. Telecomm. Union, CH-Geneva, 2017

[9] ITU-T Contr. SG12-C.300: P.SAMD Update of Ongoing Work, Source: Deutsche Telekom AG, Int. Telecomm. Union, CH-Geneva, Nov. 2018