

# Von "end-to-end" zu "mouth-to-ear"

## Zur instrumentellen Bewertung von Sprachqualität in IP-Netzen

Jens Berger, T-Nova Deutsche Telekom Innovationsgesellschaft mbH, Berkom, Berlin, jens.berger@telekom.de

### 1. Zusammenfassung und Einleitung

Die steigende Zahl an Telekommunikationsdiensten und deren Anbieter erfordert effiziente Methoden zur sicheren Bewertung der Qualität übertragener Sprache. Die sicherste und anerkannteste Methode ist immer noch die Qualitätsbewertung durch Versuchspersonen in Hör- oder Konversationstests.

In konventionellen Telefonverbindungen sind nahezu alle nicht-linearen Signalverarbeitungs-komponenten – die Einfluß auf die Sprachqualität haben – im Telekommunikationsnetz integriert. Das "klassische Endgerät" steuert in erster Linie lineare Verzerrungen an der akustischen Nutzerschnittstelle bei. Die auftretende Variationsbreite ist hier durch enge Parametervorgaben und gleichartige Schnittstellengestaltung (Handapparat) begrenzt (Bild 1).

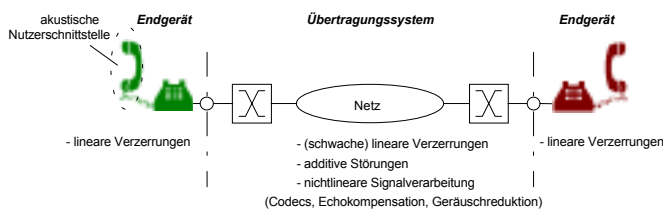


Bild 1: konventionelle Telefonverbindung

Speziell bei Sprachübertragung in IP-Netzen verschwimmen jedoch die Grenzen zwischen Netz und Endgerät. Zum einen werden Komponenten klassischer Netze wie z.B. Sprachcodierverfahren bereits in die Endgeräte integriert, zum anderen ist die Vielfalt der möglichen Endgeräte, verglichen mit dem konventionellen Handapparat, wesentlich größer. Nutzbar sind neben Handapparaten auch *headsets* oder Freisprecheinrichtungen. Diese benutzten Endgeräte und Übertragungsverfahren sind auch nicht mehr auf die Übertragung von Telefonbandsignalen begrenzt.

Eine Qualitätsbestimmung, die lediglich die Übertragungskomponenten zwischen den elektrischen Netzzugängen berücksichtigt reicht hier nicht mehr aus, um die tatsächlich von den Gesprächspartnern empfundene Sprachqualität abzubilden. Abgesehen davon, dass diese elektrischen Schnittstellen kaum noch zu definieren und zugänglich sind, ist es auch erforderlich, die Endgeräte einschließlich ihrer akustischen Schnittstellen in die Bewertung einzubeziehen.

### 2. Instrumentelle ("objektive") Beurteilung der Sprachqualität

Zur effizienten Bewertung großer Mengen von Sprachmaterial werden zunehmend instrumentelle Bewertungsverfahren eingesetzt, die eine Abschätzung der Sprachqualität anhand messbarer physikalischer Parameter der Sprachsignale vornehmen. Der Großteil der bekannten Verfahren speist dazu eine hochqualitative Sprachprobe in die Telefonverbindung ein und zeichnet das übertragene Signal an der Senke wieder auf. Die Qualitätsbewertung basiert auf einem Vergleich von Eigenschaften des ungestörten Referenz-Quellsignals und des aufgezeichneten und möglicherweise gestörten Sprachsignals. Die Signale werden dazu i.a. durch ein Gehörmodell auf eine interne Repräsentationsebene transformiert, auf dieser werden dann Unterschiede zwischen beiden Signalrepräsentationen berechnet und bewertet.

Insbesondere für die Bewertung von Sprachübertragungssystemen in realen Anwendungen können solche Verfahren genutzt werden, wenn durch vorhergehende Untersuchungen der Nachweis erbracht wurde, dass die Messverfahren die auftretenden Störungen sicher beurteilen können.

Die Schnittstellen zu diesen Messverfahren bilden immer die elektrischen Netzzugänge. Die Endgeräte sind in den Verfahren nur als ideale Modelle integriert. Die Sprachqualität wird derart prädiert, wie sie von Versuchspersonen in einer Telefonsituation unter Benutzung eines konventionell gestalteten Handapparates wahrgenommen würde (Bild 2).

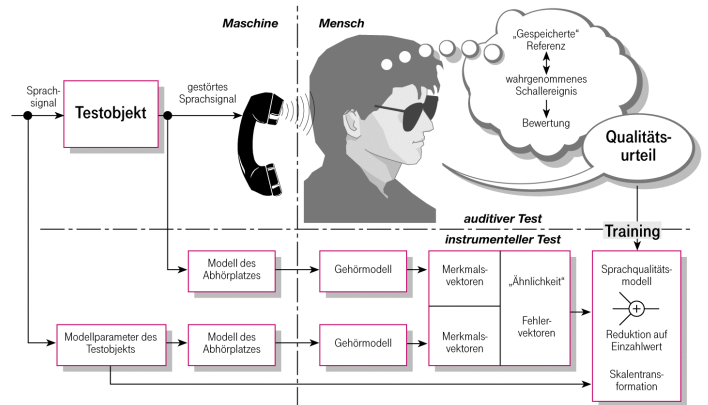


Bild 2: Struktur instrumenteller Sprachqualitätsmessverfahren

Es existieren mehrere solcher Ansätze zur Bewertung schmalbandiger Sprache, die für bekannte "netz-typische" Störungen (z.B. hervorgerufen durch Sprachcodierer, Paketverluste oder *temporal clipping*) zuverlässig die Sprachübertragungsqualität abschätzen können.

Dabei ist es zunächst unerheblich, ob diese Störungen tatsächlich im Netz entstehen oder in der Endeinrichtung beim Nutzer. Problematisch ist nur, dass die Endeinrichtungen über kein elektrisches Messinterface unmittelbar nach bzw. vor der akustischen Schnittstelle verfügen. Selbst wenn dieses existieren würde, blieben noch die Einflüsse der Schnittstelle selbst auf die Sprachqualität unberücksichtigt.

### 3. Auditive Sprachqualitätsuntersuchungen

Um den Einfluss dieser akustischen Schnittstelle im Zusammenspiel mit anderen Störungen z.B. durch Sprachcodierer oder Frequenzbandbegrenzungen zu untersuchen, wurden zunächst einige auditive Sprachqualitätsuntersuchungen mit Versuchspersonen durchgeführt. Es handelte sich hierbei ausschließlich um sogenannte *Listening-Only-Tests*, mit denen die Einwegübertragungsqualität bestimmt werden kann. Einflüsse der Signallaufzeit und des Gegensprechverhaltens bleiben hier unberücksichtigt.

Die Versuchsabläufe entsprachen den geltenden internationalen Normen im Telekommunikationsbereich, 24 untrainierte Versuchspersonen bewerteten dargebotene und sich nicht wiederholende Sprachbeispiele von 8 s Dauer und verschiedenen Sprechern und Sprecherinnen auf einer 5-stufigen Qualitätsskala, die mit den Kategorien "5: ausgezeichnet", "4: gut", "3: ordentlich", "2: dürftig", "1: schlecht" besetzt war. Da für jede zu testende Übertragungsbedingung vier verschiedene Sprachbeispiele benutzt wurden, konnten am Ende eines jeden Experiments 96 Einzelurteile zu einem *Mean Opinion Score* (MOS) zusammengefasst werden.

Zunächst wurden Aufzeichnungen von Sprachbeispielen mittel eines Kunstkopfes mit einem künstlichen Ohr Typ 3.4 nach [ITU-P57] durchgeführt. Die Sprachbeispiele waren auf einen Bereich von 200...3500Hz band-begrenzt und teilweise durch vorgeschaltete Sprachcodier verzerrt. Als akustische Schnittstellen wurden zwei konventionell gestaltete Handapparate mit unterschiedlichen Wandlern sowie eine geschlossene dynamische Kopfhörerkapsel am linken Ohr des Kunstkopfes fixiert.

Die Sprachbeispiele wurden nach entsprechender Entzerrung der Aufnahmetechnik den Versuchspersonen im Hörtest zur Bewertung angeboten. Selbst bei dieser vergleichsweise geringen Streubreite in der Gestaltung der akustischen Schnittstelle, zeigen sich vor allem im höher-qualitativen Bereich deutliche Qualitätsunterschiede (Bild 3).

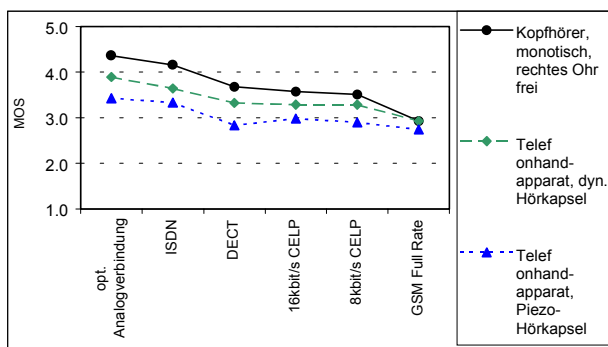


Bild 3: Sprachqualität bei verschiedenen akustischen Schnittstellen und identischen Übertragungsbedingungen

Diese Qualitätsunterschiede sind im wesentlichen durch den Verlust tieffrequenter Signalanteile durch die leck-behaftete Ankopplung der Handapparate an die Pinna zu erklären. Der Piezo-Wandler mit der höchsten akustischen Impedanz ist am leck-empfindlichsten, er zeigt auch den stärksten Qualitätsverlust.

Bei starken nichtlinearen Störungen im Signal (z.B. GSM-Sprachcodierung) überlagern diese zunehmend die Beeinträchtigungen durch die Frequenzbandeinengungen.

#### 4. Instrumentelle Sprachqualitätsmessungen einschließlich der akustischen Schnittstellen

Die Grundidee ist, einem instrumentellen Sprachqualitätsmeßverfahren, die Sprachsignale zur Verfügung zu stellen, die auch im Hörtest bewertet werden. Dies sind in den hier aufgeführten Beispielen die Aufnahmen am künstlichen Ohr. Zunächst ist zu beachten, dass die akustische Schnittstelle bereits Bestandteil der "Übertragungsstrecke" ist und somit das im Messverfahren integrierte Handapparat-Modell überbrückt werden muß. Die weitere Verarbeitung im instrumentellen Meßverfahren geht von Signalen aus, wie sie am *Ear-Referenz-Point (ERP)* vorliegen. Daher ist es zusätzlich notwendig, die mit dem Kunstkopf aufgezeichneten Signale mit einer entsprechenden Entzerrung auf diesen ERP zu transformieren.

Wesentlich problematischer ist die qualitätsgerechte Erfassung stärker linearer Verzerrungen, wie sie an der akustischen Schnittstelle auftreten. Im allgemeinen werden lineare Verzerrungen in instrumentellen Sprachqualitätsmaßen vor der Qualitätsschätzung nahezu vollständig eliminiert. Dies trifft auch für den aktuellen ITU-T Standard für instrumentelle Sprachqualitätsmessungen zu [ITU-P862]. Hier werden alle linearen Verzerrungen bis zu  $\pm 20\text{dB}$  vollständig ausgeglichen. Dies führt bei Messungen an den elektrischen Netzzugängen kaum zu Fehlern, da hier nur schwache lineare Verzerrungen zu erwarten sind.

Das instrumentelle Sprachqualitätsmessverfahren TOSQA [Berger98] wurde ursprünglich für die Bewertung von Schmalbandsprache ohne die Berücksichtigung akustischer Eigenschaften der Terminals entwickelt. Die Erweiterung des Modells erlaubt inzwischen auch die Bewertung von Breitbandsprache, es können ebenfalls Ergebnisse unter Einbeziehung der Terminals vorgestellt werden. Dazu wurden zum einen das interne Handapparat-Modell überbrückt und eine Entzerrung von Kunstkopf-Signalen integriert. Der wichtigste Schritt war jedoch die Entwicklung eines qualitäts-gerechten und begrenzten Ausgleichs linearer Verzerrungen. Dies ermöglicht, diejenigen Anteile linearer Verzerrung auszugleichen, die kaum Einfluss auf das Qualitätsurteil haben zum anderen aber störende Verzerrungen in das Qualitätsurteil einfließen zu lassen. Bild 4 zeigt die Ergebnisse des auditiven Tests (vgl. Bild 3) im Vergleich zu den Messwerten des instrumentellen Messverfahrens TOSQA2001.

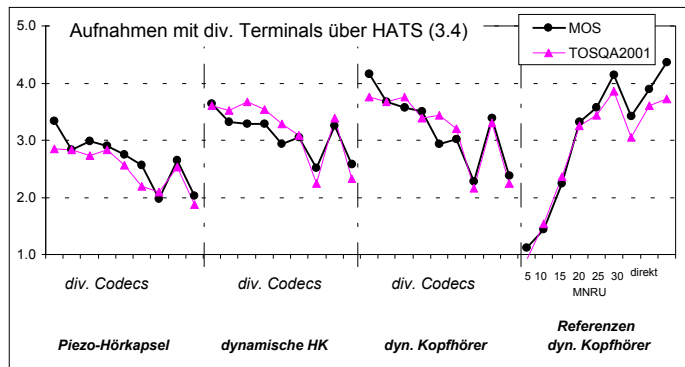


Bild 4: Vergleich von Sprachqualitätswerten aus dem auditiven Test mit instrumentell gewonnenen Ergebnissen

Eine wesentlich größere Variabilität an Endgeräten wird sich bei Verwendung von Breitbandübertragungssystemen einstellen. Bild 5 zeigt die Ergebnisse bei Bewertung verschiedener akustischer Schnittstellen, die nur z.T. breit-bandfähig sind. Die aufgezeichneten Signale waren bereits im Vorfeld band-begrenzt (s. Information innerhalb des Diagramms). Wie erwartet, fällt die Qualität eines Telefonhandapparates gegenüber einem breitbandig angesteuerten Kopfhörer deutlich ab. Aber auch bei identischer Bandbegrenzung bietet der Kopfhörer aufgrund seiner besseren Ankopplung an die Pinna höhere Qualität. Einen Sonderfall in dieser Bewertung stellen die Freisprecher dar. Zum einen wurde eine kleine HiFi-Box als Referenz gewählt, zum anderen ein preiswertes Gerät aus dem Endgerätemarkt. Auch für dieses breite Feld an Geräten und Qualitätswerten kann das instrumentelle Verfahren TOSQA2001 sichere Qualitätsschätzungen liefern.

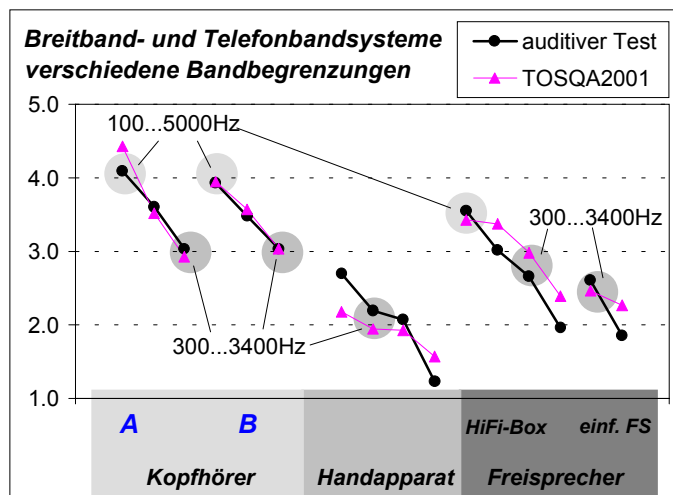


Bild 5: Sprachqualität bei Breitband- und Schmalbandsignalen

#### Literatur:

Berger, J. (1998). *Instrumentelle Verfahren zur Qualitätsschätzung - Modelle auditiver Tests*, Dissertation, Christian-Albrechts-Universität, D-Kiel.

ITU-T Rec.P57. *Artificial ears*. International Telecommunication Union, CH-Geneva.

ITU-T Rec.P862. *Perceptual Evaluation of Speech Quality (PESQ), an Objective Method for end-to-end Speech Quality Assessment of Narrow-band Telephone Networks and Speech Codecs*. International Telecommunication Union, CH-Geneva.