

# Testverfahren zur Sprachqualitätsbeurteilung von VoIP-Systemen

H.W. Gierlich, F. Kettler, E. Diedrich\*

HEAD acoustics GmbH, Ebertstraße 30a, 52134 Herzogenrath

\* T-Nova Deutsche Telekom Innovationsgesellschaft mbH Berkorn

## 1. Einleitung

In Voice over IP Systemen wird die Sprachqualität durch eine Vielzahl von Parametern beeinträchtigt:

Die Laufzeiten, die weder vorhersagbar noch konstant innerhalb einer Verbindung sind, haben einen signifikanten Einfluß auf die Konversationsqualität. Hierfür gibt es im wesentlichen zwei Gründe: Einerseits wird die Interaktionsfähigkeit zwischen zwei Teilnehmern bei zunehmender Laufzeit immer mehr erschwert. Andererseits erhöht sich bei zunehmender Laufzeit auch die Empfindlichkeit des Sprechers bzgl. der Wahrnehmung von Echostörungen.

Sprachgesteuerte Verstärkungsregelungen und Sprachschalter können an vielen Stellen der Verbindung auftreten: im Endgerät sowie an verschiedensten Stellen der Verbindung. Insbesondere kann die Kaskadierung verschiedener Algorithmen zur Sprachdetektion auch hier zu zusätzlichen signifikanten Qualitätseinbußen führen, wobei nicht nur die Sprachqualität sondern sogar die Sprachverständlichkeit beeinträchtigt werden kann. Sowohl Echo- als auch Schaltprobleme können während Einzel- und Doppelsprechen auftreten und sind perceptiv in diesen verschiedenen Situationen unterschiedlich zu bewerten.

Einen weiteren wesentlichen Einfluß auf die Qualität einer Verbindung hat die Übertragungsqualität des Hintergrundgeräusches, die oft ignoriert wird. Auch hier spielt die Kaskadierung von Signalverarbeitungskomponenten zwischen Terminal und Netz eine entscheidende Rolle: Während typischerweise im Terminal bereits Algorithmen zur Störgeräuschreduktion eingesetzt werden, versuchen zusätzlich Netzkomponenten Segmente, in denen lediglich Störgeräusch und keine Sprache übertragen wird, zur Reduktion der benötigten Bandbreite herauszuschneiden und durch „Comfort Noise“ zu ersetzen. Dies führt für den Teilnehmer u.U. zur Wahrnehmung eines modulierten, nicht kontinuierlichen Hintergrundgeräusches. Auch hier ist wiederum die Einzel- und Gegensprechsituation zu betrachten, da sowohl Signalverarbeitung als auch die Perception der Störung unterschiedlich ist.

Unabhängig von den vorab beschriebenen Effekten können zusätzliche Störungen durch Paketverluste, Delay und Jitter auftreten, die sowohl Segmente, in denen Sprache übertragen wird, betreffen als auch Segmente, in denen Hintergrundgeräusch übertragen wird. Auch diese Effekte wiederum sind unterschiedlich zu bewerten, abhängig von der Qualität der Rekonstruktionsalgorithmen sowie der Konversationsituation, in denen diese Effekte auftreten.

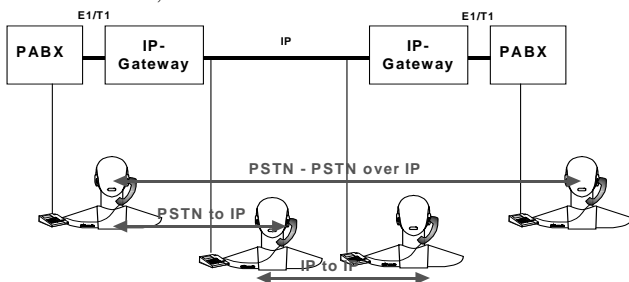


Abb. 1: Typische Szenarien in VoIP-Netzwerken

Ein wesentliches Merkmal von Voice over IP Konfigurationen im Vergleich zu herkömmlichen Telekommunikationsszenarien ist die Tatsache, daß eine eindeutige Unterscheidung zwischen Netz- und Terminalkomponenten nicht mehr möglich ist. Dies bedeutet, daß meßtechnisch immer die komplette Konfiguration betrachtet werden sollte, d.h. vom akustischen Zugangspunkt (Terminal) bis hin zu einem definierten Punkt im Netz bzw. im Idealfalle bis zum anderen akustischen Zugangspunkt des B-Teilnehmers (vergleiche Abb. 1). Darüber hinaus muß davon ausgegangen werden, daß der klassische Handapparat in vielen Fällen durch Freisprecheinrichtungen bzw. Hörsprechgarnituren ersetzt wird.

## 2. Verfahren- und Analysemöglichkeiten zur Sprachqualitätsbestimmung

Für die Beeinträchtigung der Konversationsqualität müssen drei Situationen betrachtet werden:

- die Hörsituation,
- die Sprechsituation sowie
- die Konversationsituation.

Die Sprachqualität wird in diesen Situationen durch unterschiedliche Parameter beeinträchtigt, entsprechend müssen die adäquaten Parameter untersucht werden. Tabelle 1 zeigt in einer Übersicht die wesentlichen objektiven Parameter, die zur Bestimmung der Sprachqualität in den verschiedenen Situationen herangezogen werden können.

Hörsituation	Sprechsituation	Konversationsituation
<ul style="list-style-type: none"> <li>- „Standard“-Parameter, z.B. Loudness Ratings</li> <li>- psychoakustisch motivierte Verfahren, z.B. PESQ, TOSQA</li> <li>- erweiterte Messungen, z.B. Schaltcharakteristik</li> <li>- Hintergrundgeräuschübertragung, z.B. Relative Approach</li> </ul>	<ul style="list-style-type: none"> <li>- Laufzeit</li> <li>- Echodämpfung</li> <li>- Echo-/Rückhörverzerrung</li> <li>- Vorschläge für psychoakustisch motivierte Testverfahren, z.B. PESQM (perceptual echo and sidetone quality measure)</li> <li>- Hintergrundgeräuschübertragung, z.B. Relative Approach</li> </ul>	<ul style="list-style-type: none"> <li>- Laufzeit</li> <li>- Echodämpfung</li> <li>- Echo-/Rückhörverzerrung</li> <li>- Schaltcharakteristika</li> <li>- Zeitvariante Echobeurteilung</li> </ul>

Tabelle 1: Objektive Parameter für die verschiedenen Konversationssituationen

### 2.1 Testverfahren für die Hörsituation

In der Hörsituation kommen zunächst die "klassischen" Parameter wie Loudness Ratings, Übertragungsfunktionsmessung/-schätzung, Verzerrungsmessung etc. zum Einsatz. In einfacher Weise können aber auch psychoakustisch motivierte Verfahren wie PESQ oder TOSQA eingesetzt werden [1], [2]. Derartige Verfahren sind insbesondere geeignet, unter verschiedenen Netzbedingungen (Packet Loss und Delay Jitter) die Beeinträchtigungen der Einwegsprachübertragungs-

qualität für die Hörsituation zu ermitteln, da derartige Verfahren einen geschätzten MOS-Wert (Mean Opinion Score) als Qualitätswert liefern. Zu beachten ist allerdings, daß die meisten Verfahren lediglich zur Sprachqualitätsbestimmung zwischen zwei elektrischen Zugangspunkten geeignet sind, die akustischen Eigenschaften des Terminals aber weitgehend unberücksichtigt bleiben. Weitergehende Informationen hierzu finden sich in [2]. Wie fast alle anderen Analysen berücksichtigen diese Verfahren allerdings lediglich das Nutzsignal Sprache, die Gesamtqualität in der Hörsituation wird aber darüber hinaus auch durch die Qualität der Hintergrundgeräuschübertragung wesentlich beeinflusst.

Eine objektive Erfassung der Qualität der Hintergrundgeräuschübertragung ist schwierig, da Variationen im Hintergrundgeräusch sowohl von der Quelle selbst als auch von der zwischengeschalteten Übertragung stammen können. Da in der Regel das Hintergrundgeräusch nicht bekannt ist, müssen zur Analyse Verfahren angewandt werden, die keine Kenntnis über das Hintergrundgeräusch zur Analyse benötigen, sogenannte referenzsignalfreie Testverfahren. Dieser Ansatz ist möglich unter der Annahme eines relativ konstanten Hintergrundgeräusches wie es teilweise in Büroumgebungen aber auch in typischen Fahrsituationen auftritt. Gute Ergebnisse konnten mit dem "Relative Approach" [3] erzielt werden. Der auditive Eindruck korreliert für die oben beschriebenen Situationen gut mit den Ergebnissen des Algorithmus. Wesentlicher Bestandteil dieses Gehörmodell basierten Verfahrens [4] ist die gehöradäquate Bewertung des Kurzzeitpegels (ca. 2 ms) im Vergleich zum Langzeitpegel (gemittelt über 2-4s) sowohl in zeitlicher Richtung als auch in barkspektralen Bereich. Abweichungen sowohl im Zeit- als auch im Frequenzbereich können in einer Art Spektrografie visualisiert werden. Abbildung 2 zeigt für ein typisches Hintergrundgeräusch eine derartige Analyse.

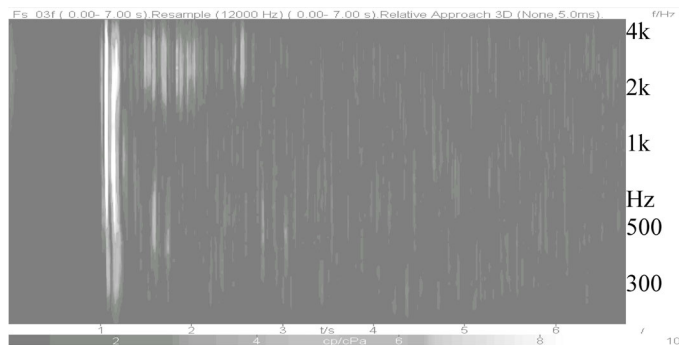


Abb. 2: Relative Approach Analyse (helle Farben repräsentieren wahrnehmbare Störungen)

Die Abbildung zeigt den Adaptionsvorgang eines Hintergrundgeräusch-Reduktionsalgorithmus bei gleichförmigem Hintergrundgeräusch (Fahrzeuginnengeräusch). Auditiv störend ist in diesem Beispiel der Einschaltvorgang (breitbandige Komponente zu Beginn des Adaptionsalgorithmus) sowie zeit- und frequenzselektive Strukturen, die sich innerhalb der ersten Sekunde auch beim Relative Approach bemerkbar machen.

## 2.2 Testverfahren für die Konversationssituation

In der Konversationssituation befinden sich die Gesprächsteilnehmer sowohl in der Hör- als auch in der Sprechsituation. Besonders problematisch sowohl seitens der Meßtechnik als auch seitens der Implementierung ist allerdings die Doppelsprechsituation in der die Sprachqualität durch zeitvariante Echostörungen, Schalten (ebenfalls zeitvariant) und Hintergrundgeräuschmodulation beeinträchtigt wird. Grenzwerte bezüglich der Echostörungen sowie der Störungen durch Schalten wurden bereits in [5] vorgestellt. Hier zeigt sich, daß zwischen der Reduktion der Echodämpfung beim Gegensprechen und beim Einzelsprechen bei gleichbleibender Echostörung ein funktionaler Zusammenhang besteht. Dies bedeutet, daß zur Erzielung einer möglichst guten Gesamtqualität die Echodämpfung während des Gegensprechens nicht mehr als 4 - 6 dB gegenüber der Echodämpfung während des Einzelsprechens abfallen sollte. Ähnliches gilt für Schalten: Um auch hier eine möglichst hohe Qualität zu gewährleisten darf der Pegelhub während des Gegensprechens nicht mehr als 6 dB

betragen (weitere Informationen siehe [5]). Echostörungen werden typischerweise während des Gegensprechens dann wahrgenommen, wenn für kurze Zeit das Gegensprechsignal einen geringen Pegel oder eine Pause aufweist. Zu diesen Zeitpunkten ist es auch sinnvoll, die Echodämpfung zu bestimmen. Ähnliches gilt für das Schalten, auch hier wird Pegelvariation des Gegensprechsignals typischerweise dann wahrgenommen, wenn der Zuhörer selber nicht spricht. Basierend auf diesen Erkenntnissen können sprachähnliche Signalkombinationen verwendet werden wie sie in Abbildung 3 dargestellt sind. In Abbildung 4 ist eine typische Analyse basierend auf einer derartigen Meßsignalkombination prinzipiell dargestellt. Aus den entsprechenden Signalabschnitten können sowohl Echodämpfung als auch Schaltcharakteristik abgelesen werden.

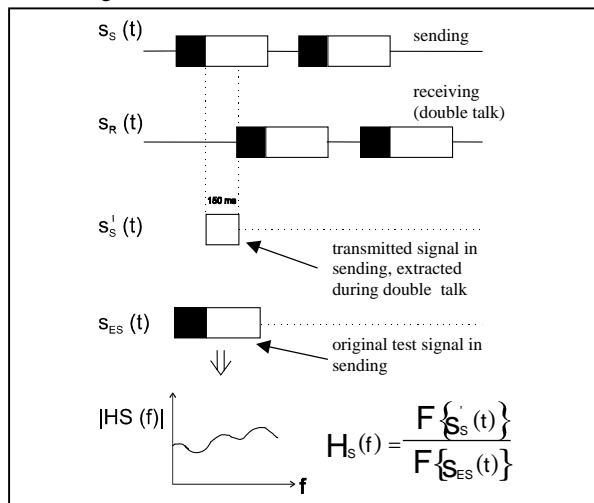


Abb. 3: Testsignale zur Untersuchung der Gegensprechsituation

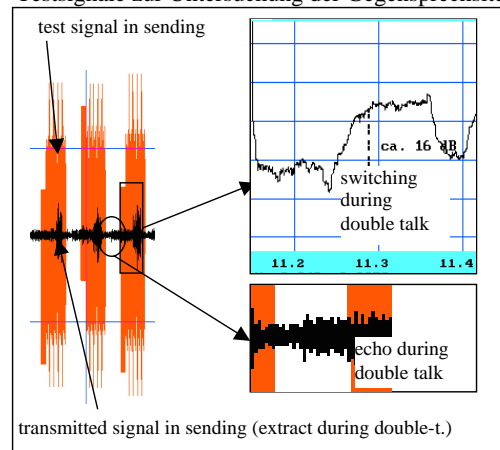


Abb. 4: Typische Meßergebnisse während Doppelsprechens

## 4. Zusammenfassung

In einer Übersicht wurden typische Probleme von VoIP-Systemen und geeignete Testverfahren zur Untersuchung der Sprachqualität in verschiedenen Hör- und Sprechsituationen vorgestellt. Für verschiedene Aspekte der Sprachqualität stehen bereits objektive Untersuchungsmethoden zur Verfügung. Ein Einzahlwert als umfassende Aussage bezüglich der Gesamtsprachqualität kann aber mit derzeit verfügbaren Methoden noch nicht abgeleitet werden.

## 5. Literatur

- [1] Berger, J., Instrumentelle Verfahren zur Sprachqualitätsschätzung - Modelle auditiver Tests, Dissertation, 1998, Shaker Verlag
- [2] ITU-T Recommendation, Draft P.862: Perceptual Evaluation of Speech Quality (PESQ)
- [3] Genuit, K., Objective Evaluation of Acoustic Quality Based on a Relative Approach, Inter-Noise'96, Liverpool, England
- [4] Sottek, R.: Modelle zur Signalverarbeitung im menschlichen Gehör, Dissertation, RWTH Aachen, 1993
- [5] F. Kettler; Gerlich, H.W.; Diedrich, E. Echo and Speech Level Variations During Double Talk Influencing Hands-free Telephones Transmission Quality, IWAENC 99, Pocono Manor, USA