

Sprachqualitätsbeurteilung von IP-Systemen

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

HEAD acoustics GmbH, Ebertstraße 30a, 52134 Herzogenrath

*T-Nova Berkom GmbH, Goslarer Ufer 35, 10589 Berlin

1. Einleitung

Im Zuge der Konvergenz von Sprachdaten in Multimediadiensten bekommt die Sprachübertragung über IP-basierte Netze immer mehr Bedeutung. Typische Endgeräte, die in diesen Netzen zu finden sein werden, sind weniger die traditionellen Handapparate-Terminals sondern eine Vielfalt neuer Designs, angefangen von laptopbasierten Kommunikationssystemen bis hin zu drahtlosen Headsets. Derartige Terminals sind bereits aus Sicht der Akustik relativ komplex, häufig muß mit preiswerten, einfachen Wandlern in akustisch ungünstiger Anordnung eine vernünftige Sprachübertragungsqualität realisiert werden. Dies kann mit rein akustischen Maßnahmen nicht gewährleistet werden. Zur Erzielung einer ausreichenden Sprachqualität sind verschiedenste Signalverarbeitungsmechanismen erforderlich.

2. IP-Terminal/Konfiguration

Betrachtet man eine Sprachübertragung über IP-Systeme aus Sicht der Sprachübertragungsqualität, so läßt sich ein typisches IP-Terminal wie in Abbildung 1 gezeigt charakterisieren.

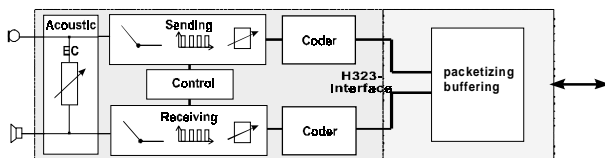


Abb. 1: Signalverarbeitung in IP-Terminals (vereinfacht)

Die typischen Komponenten der Signalverarbeitung sind prinzipiell bekannt: Akustische Echokompensation, sprachgesteuerte Verstärkungsregelungen, Teilbandsignalverarbeitungen, Kodierung, Geräuschreduktion, Paketisierung aber auch VAD (Voice Activity Detection) sind Signalverarbeitungsverfahren, die letztlich die Sprachübertragungsqualität von A nach B bestimmen. Darüber hinaus finden sich aber noch weitere Parameter, die in der traditionellen Telefonometrie unbekannt waren und die Sprachübertragungsqualität erheblich beeinflussen:

- Paketverlust
- variables Delay (innerhalb einer Verbindung)
- Jitter Probleme

Alle oben genannte Signalverarbeitungsverfahren bedingen, daß Sprachübertragung über IP-Netze in hohem Maße durch nichtlineare und zeitvariante Systeme beeinträchtigt wird. Demzufolge sind die wichtigsten Parameter, die die Konversations Sprachqualität beeinflussen:

- Delay und Clipping
- Qualität der Hintergrundgeräuschübertragung
- Gegensprechfähigkeit
- Echostörung unter Einzel- und Doppelsprechbedingungen
- Sprachklangqualität
- (Sprach-) Klangqualität

Die klassischen, aus der Telefonometrie bekannten Parameter wie Loudness Ratings und Übertragungsfunktion haben relativ zu den oben beschriebenen Störungen nur noch eine untergeordnete Bedeutung, solange sie in einem akzeptablen Rahmen liegen.

3. Messungen an IP-Konfigurationen

In den vergangenen Jahren wurden, basierend auf Grundlagenuntersuchungen die an Freisprechtelefonen durchgeführt wurden, Meßverfahren zur Qualifizierung von nicht LTI-Systemen entwickelt. Diese Testverfahren sind prinzipiell auf alle nicht LTI-Systeme anwendbar, da die sprachqualitätsbeeinflussenden Faktoren durchaus vergleichbar sind.

An 2 IP-Konfigurationen, die in Abbildung 2 und 3 dargestellt sind, wurden verschiedene Messungen zur objektiven Bestimmung der Sprachübertragungsqualität durchgeführt. In beiden Konfigurationen traten weder Paketverluste noch Jitterprobleme auf.

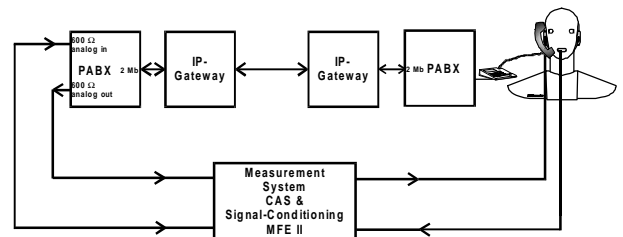


Abb. 2: IP-Konfigurationen 1

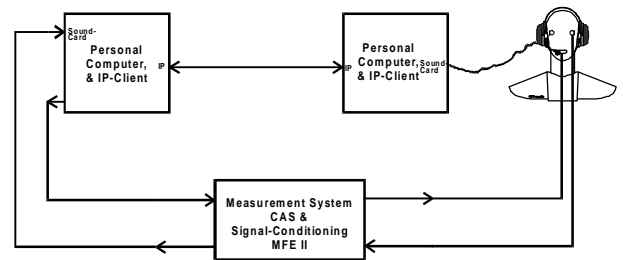


Abb. 3: IP-Testkonfigurationen 2

Anhand einiger Beispiele werden die Ergebnisse der Messungen im folgenden erläutert. Die verwendeten Meßverfahren wurden basierend auf auditiven Tests entwickelt, die u. a. in [1] und [2] beschrieben sind.

3.1 Laufzeit und Echodämpfung

Die zunächst wichtigsten Einflußfaktoren von IP-Konfigurationen sind - neben den traditionellen Parametern wie Loudness Ratings und Übertragungsfunktionen - Laufzeit und Echodämpfung. Die Analyse der Konfiguration 1 ergab eine Laufzeit von ca. 60 ms bei einer ausreichenden Echodämpfung von >40 dB. Die Laufzeit der Konfiguration 2 war variabel, abhängig vom Verbindungsaufbau zwischen 480 und 550 ms, die Echodämpfung betrug lediglich ca. 21dB, dies ist absolut unzureichend, die Anforderung nach ITU-T Empfehlung G. 131 wäre mindestens 56 dB Echodämpfung.

3.2 Hintergrundgeräuschübertragung in Senderichtung

Für diesen Test wird ein langsam ansteigendes, rauschähnliches Signal verwendet, mit dem Schalten, Nichtlinearitäten etc. bei Hintergrundgeräuschübertragung festgestellt werden können.

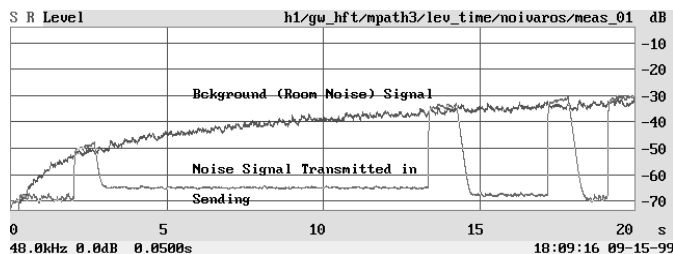


Abb. 4: Hintergrundgeräuschübertragung in Senderichtung für Konfiguration 2

An diesem Bild läßt sich deutlich, die auch auditiv bemängelte, schlechte Qualität der Hintergrundgeräuschübertragung erkennen: unabhängig vom Signalpegel, wird zu verschiedenen Zeitpunkten scheinbar willkürlich das Hintergrundgeräusch an- bzw. abgeschaltet. Dies führt zu einer äußerst unangenehmen Übertragungscharakteristik für das Hintergrundgeräusch.

3.3 Pegelabhängige Nutzsinalübertragung in Empfangsrichtung

Hierzu wurde ein spezielles Testsignal verwendet, das aus langsam ansteigenden, stimmhaften Komponenten von Sprache besteht. Abbildung 5 zeigt das Meßergebnis für Konfiguration 2. Es ist deutlich zu erkennen, daß für fast alle Signalpegel eine erhebliche Kompression stattfindet, d. h. unabhängig vom Leitungspegel wird das Nutzsinal fast gleich laut übertragen. Dieses in bestimmten Pegelbereichen sinnvolle Verhalten führt - angewandt auf den gesamten Pegelbereich - zu einer extremen Beeinträchtigung der Einwegsprachqualität bei dem Hintergrundgeräusch, da während fortlaufend gesprochene Sprache das Hintergrundgeräusch in gleicher Weise mit komprimiert wird.

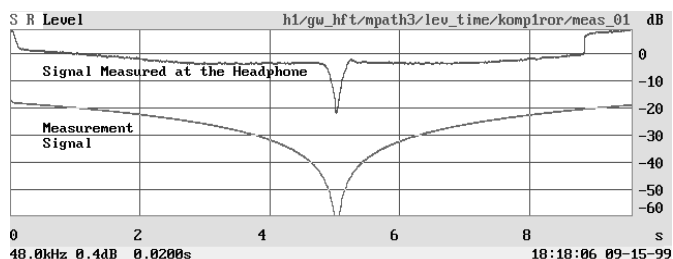


Abb. 5: Signalübertragung in Empfangsrichtung für Konfiguration 2

3.4 Gegenseprechfähigkeit

Zur Bestimmung der Gegenseprechfähigkeit wurde ein spezielles Testsignal entwickelt, das aus zeitversetzten, unkorrelierten, langsam ansteigenden bzw. abfallenden Composite Source Signalen (sprachähnlich) besteht [3]. Mit Hilfe dieses Testsignals lassen sich sowohl Schalten als auch Echo während der Gegenseprechphase untersuchen. Für ein gutes Gegenseprechverhalten sollte der Pegelhub auch in einer Doppelsprechphase < 6 dB und die Echodämpfung mindestens einen Wert von 37 dB bei einer Laufzeit von 100 ms erreichen. Die Grundlagenuntersuchungen hierzu sind in [4], [5] veröffentlicht. Während für Konfiguration 1 eine ausreichend hohe Echodämpfung während des Gegenseprechens mit > 40 dB erzielt wurde, ist die Echodämpfung für Konfiguration 2 auch im Doppelsprechbetrieb entweder nur 20,8 dB (d.h. es werden keinerlei echounterdrückenden Maßnahmen eingesetzt) oder aber die Senderichtung wird um mehr als 20 dB bedämpft was bezüglich des Schaltverhaltens nicht akzeptabel ist.

Die Schaltcharakteristik für Konfiguration 1 während des Gegenseprechens kann beispielhaft anhand Abbildung 6 dargestellt werden, hier ist das Zeitsignal zu sehen, das in Senderichtung bei einem Pegel von -23 dBPa aufgezeichnet wurde. Der Pegel des Gegenseprechsignals in Empfangsrichtung betrug -8,3 dBV.

Dargestellt ist das Testsignal sowie die Reaktion des Systems in Senderichtung auf dieses Testsignal. Es ist deutlich zu erkennen, daß selbst bei dieser ungünstigen Kombination die Senderichtung bereits 100 ms nach dem Doppelsprechvorgang wieder voll aktiviert ist. Da für Signalpegel in Sende- und Empfangsrichtung, die in etwa den Standardpegel entsprechen, keinerlei Beeinträchtigung festgestellt werden konnte, kann davon ausgegangen werden, daß dieses System Gegenseprechfähigkeit in vollem Umfang gewährleistet.

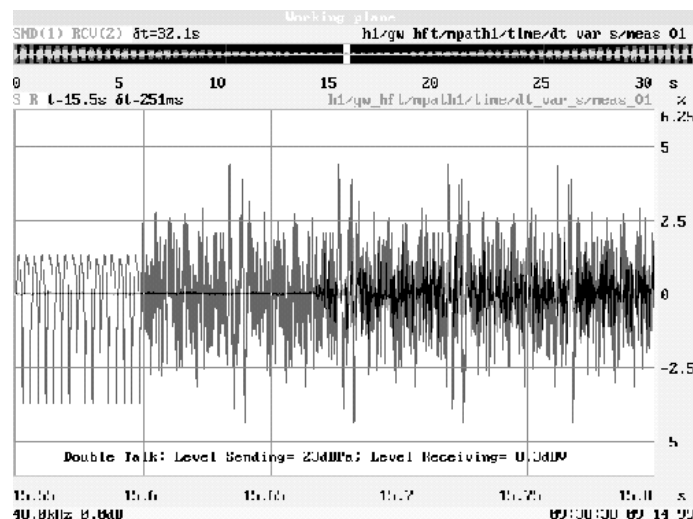


Abb. 6: Schaltcharakteristik

4. Zusammenfassung

Die vorgestellten instrumentellen Untersuchungen an zwei IP Konfigurationen zeigen, daß unter guten Randbedingungen (keine Paketverluste und Jitter) IP-Systeme durchaus akzeptable Sprachqualität bieten können. Bei schlechtem Design allerdings ist die Sprachqualität absolut inakzeptabel (Konfiguration 2: hohe Laufzeit, geringe Echodämpfung, starke Kompandierungs- und Clippingeffekte). Die Untersuchungen zeigen weiterhin, daß die Gesamtsprachqualität wesentlich durch die Übertragungseigenschaften des Terminals mitbestimmt werden.

5. Literatur

- [1] Gierlich, H.W.; Kettler, F.; Krebber, W., Subjective Evaluation Procedures for Hands-Free Telephones - Double Talk Performance-, ITU-T SG 12 Meeting, 07.-18.04.1997, Genf, COM 12-05E
- [2] Gierlich, H.W.; Krebber, W.; Kettler, F., Subjective Evaluation of Hands-Free Telephones Using Conversational Tests, Specific Double Talk Tests and Listening Only Tests, ITU-T SG 12 Meeting, 07.-18.04.1997, Genf, COM 12-06E
- [3] ITU-T Recommendation P.501, ITU-T Series P Recommendation, October 1996
- [4] H.W. Gierlich; F. Kettler; E. Diedrich, Sprachqualitätsbestimmende Parameter für Freisprecheinrichtungen in der Gegenseprechsituation, DAGA'98, 23.-27.03.1998, Zürich
- [5] Gierlich, H.W.; Kettler, F.; Diedrich, E., <Speech Quality Evaluation of Hands-Free Telephones During Double Talk: New Evaluation Methodologies, Eusipco'98, 08.-11.09.1998, Rhodos, Greece