

Standardisierung instrumenteller Sprachqualitätsmessverfahren Berücksichtigung der akustischen Schnittstellen in der Sprachtelefonie

Jens Berger; John G. Beerends*; Antony W. Rix**

Deutsche Telekom T-Systems Nova GmbH Berkom; *TNO Telecom (NL); **Psytechnics Ltd. (UK)

1. Einleitung

Bis vor einigen Jahren war ein elektroakustisch eng definierter Handapparat die Schnittstelle zum ‚Fernsprechnetz‘. Aufgrund dieser engen Grenzen, genügte eine Betrachtung der qualitativen Veränderung zwischen den beiden Netzzugangspunkten und die Berücksichtigung eines modellhaften Handapparates, um eine Abschätzung der Gesamtsprachqualität der zu testenden Verbindung vornehmen zu können.

Die rasant zunehmende Komplexität der Netzstrukturen und eine hohe Vielfalt möglicher Endgeräte wird den klassischen Handapparat verdrängen und auch der Begriff des Fernsprechnetzes passt immer weniger zu den tatsächlichen Gegebenheiten. Die Endgeräte übernehmen Funktionen, die Komponenten klassischer Netze waren (z.B. Sprachcodierverfahren bei Mobilfunk oder IP-Phones). Auch verändertes Telefonieverhalten erfordert Maßnahmen im Endgerät (z.B. Echokompensator, Geräusch- und Nachhallreduktion). Soll nun die Qualität der Gesamtverbindung beurteilt werden, müssen die Endgeräte einschließlich ihrer akustischen Schnittstellen in die Bewertung einbezogen werden.

Die ITU-T stellt sich mit den derzeit laufenden Aktivitäten diesen Anforderungen. Unter dem Arbeitstitel P.AAM (Acoustic Assessment Model) wurde zum Wettbewerb um einen neuen Standard mit der Möglichkeit der instrumentellen (‚objektiven‘) Messung der Sprachübertragungsqualität an der akustischen Schnittstelle aufgerufen.

2. Aktuelle Standardisierung in der ITU-T

Innerhalb der ITU-T werden seit vielen Jahren instrumentelle Messverfahren standardisiert, die anhand der Analyse von Eigenschaften der Sprachsignale Abschätzungen der Qualität vornehmen. Grundsätzlich lässt sich sagen, dass die bisher standardisierten Modelle auf die Bewertung der ‚Einwegsprachübertragungsqualität‘ ausgerichtet sind. Sie schätzen die Sprachübertragungsqualität anhand eines Vergleiches des übertragenen (und ggf. gestörten) Sprachsignals zu einem hochqualitativen Referenzsignal (dem Eingangssignal der Verbindung) ab. Solche Modelle benötigen daher ‚Probeverbindungen‘ in denen Sprachbeispiele übertragen und nach dem Aufzeichnen ausgewertet werden.¹

Die Abschätzung der Sprachqualität erfolgt auf einer Skala die dem Absolute Category Rating (ACR) zur Bewertung der Gesamtsprachqualität (Listening Quality, LQ) angelehnt ist [1]. Eine Adaption dieser Modelle an ein Degredation Category Rating (DCR) oder eine Bestimmung der Höranstrengung (Listening Effort, LE) ist teilweise möglich. Diese Modelle bewerten Veränderungen im Sprachsignal, die während der Wiedergabe störend wahrgenommen werden. Der bekannteste Vertreter solcher Modelle ist sicherlich das Verfahren ITU-T P.861 ‚PSQM‘ [2], es wurde 2001 durch die neue Empfehlung ITU-T P.862 ‚PESQ‘ ersetzt.[3][4][5] Auf dem Markt existieren jedoch viele andere Verfahren mit gleicher Zielrichtung, z.B. PACE, TOSQA, PAMS, Squad und andere.

¹ Zusätzlich ist zu erwähnen, dass sich im Moment sich mehrere Firmen um ein standardisiertes Messverfahren bemühen, das ohne Kenntnis des Eingangssignals eine Abschätzung der Sprachqualität erlaubt (P.SEAM, Single Ended Assessment Model). Auch existieren bereits Vorschläge zur Bewertung der ‚talking quality‘, also der Störungen, die während des Sprechens selbst wahrgenommen werden.

3. Messungen an der akustischen Schnittstelle

Alle diese bekannten Modelle sind für die Bewertung von Sprachsignalen ausgelegt, die an der elektrischen Netzschnittstelle eingespeist bzw. aufgezeichnet wurden. Sie arbeiten unter Annahme eines konventionellen Handapparates der modellhaft im Verfahren integriert ist. Für die hier zu erwartenden Störungen liefern die Modelle gute bis sehr gute Qualitätsvorhersagen.

Ein Einsatz dieser Messverfahren zur Bewertung von Aufzeichnungen an der akustischen Schnittstelle z.B. mit einem Ohrsimulator oder ‚Kunstkopf‘ ist nicht möglich. Dies hängt zum einen mit dem bereits im Modell enthaltenen Endgerät zusammen, bei Auswertung der akustischen Aufzeichnung würde dieses modellhafte Endgerät in Serie mit dem tatsächlichen liegen und vor allem treten an der akustischen Schnittstelle Störungen und Effekte auf, auf welche die bisherigen Modelle nicht ausgerichtet worden sind.

Eine Erweiterung der Modelle sollte Messungen an der akustischen Schnittstelle an der Sendeseite oder der Empfangsseite bzw. an beiden Seiten ermöglichen. Damit können vier Kombinationen von Messaufbauten abgedeckt werden:

- ‚Elektrisch‘ – ‚Elektrisch‘ (bereits mit ITU-T P.862 möglich)
- ‚Elektrisch‘ – ‚Akustisch‘
- ‚Akustisch‘ – ‚Elektrisch‘
- ‚Akustisch‘ – ‚Akustisch‘

Derzeit existieren einige nicht standardisierte Verfahren, die zumindest in einigen akustischen Messaufbauten einsetzbar sind. Dies sind z.B. eine Erweiterung des Verfahrens PAMS. [6][7] Das Verfahren TOSQA2001 bietet sowohl eine Erweiterung für Messungen an der akustischen Schnittstelle als auch die für Bewertung von Breitbandsprache. [8][9] Dieses Verfahren wurde während des 1st und 2nd ETSI VoIP Speech Quality Test Event erfolgreich eingesetzt. [10][11]

Die Anforderungen der ITU-T an das neue Verfahren P.AAM sind vielschichtig. Zum einen muss es im ‚konventionellen‘ Messaufbau bei Verwendung elektrischer Schnittstellen vergleichbare oder bessere Ergebnisse liefern als der derzeitige Standard ITU-T P.862. Zum anderen sind Messaufgaben an der akustischen Schnittstelle zu erfüllen. Dies sind z.B. Bewertung von Endgeräten einschließlich Freisprechanwendungen, reale akustische Geräusche auf der Empfangs- und Sendeseite und auch die damit verbundenen Einflüsse der Raumakustik in den Messräumen.

Um dies zu erreichen, müssen die bisherigen Ansätze um wesentliche Aspekte erweitert werden.

1. Überbrückung des Handapparate-Modells bei Bewertung von akustischen Sprachaufnahmen
2. Einsatz breitbandiger EingangsfILTER zur Bewertung (breitbandiger) Geräusche auf der Empfangsseite
3. Erfassung und Bewertung typischer Störungen und Effekte von Endgeräten und der akustischen Schnittstelle (z.B. starke Verzerrungen im Frequenzgang)
4. Modellierung (einfacher) binauraler Effekte
5. Berücksichtigung von geringeren Schallpegelwerten bei Freisprechen im Vergleich zum Handapparat
6. Berücksichtigung des Einflusses der Raumakustik auf die Sprache (z.B. Freisprechen in halligen Räumen)

4. Aktueller Stand P.AAM und vorläufige Ergebnisse

Nach Aufruf für Vorschläge für eine instrumentelles Messverfahren zur Bewertung von Sprache an der akustischen Schnittstelle von Telefonverbindungen signalisierten die drei Firmen ihre Bereitschaft an einem Wettbewerb, die durch die Autoren dieses Beitrages vertreten werden. Basis des neuen Verfahrens P.AAM soll in jedem Fall der Kern des derzeitigen Standards P.862 sein. Aufgrund der komplexen Anforderungen entschlossen sich die drei Wettbewerber zu einer Kooperation und werden ein gemeinsames Verfahren vorschlagen.

Die Ergebnisse der drei Einzelverfahren, die von den jeweiligen Partnern beigesteuert werden, wurden bereits innerhalb der ITU-T veröffentlicht. [12]

Alle drei Verfahren basieren auf dem Kern des derzeitigen Standards P.862, es wurden neben den notwendigen Erweiterungen für die akustischen Messungen auch zahlreiche Optimierungen und Verbesserungen des P.862 – Kerns vorgenommen. Bereits für den ‚konventionellen‘ Messaufbau, die Messung zwischen zwei elektrischen Schnittstellen, übertrifft jedes der drei Verfahren die Sicherheit der Ergebnisse, die mit ‚PESQ‘ nach ITU-T P.862 für eine vorgegebene Zahl von Datenbanken erzielt wurden. Die Erhöhung der Korrelation zu den Ergebnissen entsprechender Hörtests betrug in kritischen Einzelfällen bis zu 10%. Über alle 19 Sprachdatenbanken konnte die Korrelation um 2% im Mittel von 92.6% auf ca. 94.7% gesteigert werden.

Aus den Ergebnissen läßt sich aber auch erkennen, dass der Fokus der einzelnen Ansätze verschieden ist. So ist nur eines dieser drei Verfahren in der Lage, akustische Geräusche auf der Empfangsseite sicher zu bewerten. Ein anderes dieser drei Verfahren zeigt ausgesprochen gute Ergebnisse bei der Beurteilung verschiedener Endgeräte mit starken nichtlinearen Verzerrungen und Freisprechern. Die Arbeiten zur Integration der wesentlichen Bestandteile in ein gemeinsames Verfahren dauern derzeit noch an und sollen Ende April 2003 abgeschlossen werden. Der derzeitige Stand läßt die schon erwähnten Verbesserungen im ‚konventionellen‘ Messaufbau an den elektrischen Schnittstellen erwarten. Für die verschiedenen akustischen Anwendungen werden für die zur Verfügung stehenden Sprachdatenbanken Korrelationen von mehr als 90% erreicht, im Mittel derzeit 92%.

Als Beispiel für die vorläufigen Resultate soll ein Beispiel einer Sprachdatenbank dienen, die einen sehr weit gesteckten Bereich von Testbedingungen umfasst. Dies sind verschiedene monaurale und binaurale Handapparate und Headsets, die mit verschiedenen Anpresskräften am künstlichen Ohr fixiert wurden. Ebenfalls sind hier verschiedene Freisprecheinrichtungen bewertet worden. Die Rohergebnisse wurden in den Skalenbereich von 1...5 transformiert, die Korrelation mit den Hörtestergebnissen beträgt hier über 92% (Bild 1).

Die Ergebnisse zeigen eine sehr gute Übereinstimmung mit den Hörtestwerten. Lediglich bei der Bewertung einiger Freisprecheinrichtungen macht sich das derzeit noch nicht integrierte binaurale Modell durch größere Abweichungen der Ergebnisse bemerkbar.

Die Arbeiten werden bis April 2003 abgeschlossen, nach der Auswertung der Ergebnisse und einer mehrmonatigen Testphase durch unabhängige Partner soll das Modell – wenn alle Anforderungen erfüllt werden – im September 2003 als neuer ITU-T Standard vorgeschlagen werden.

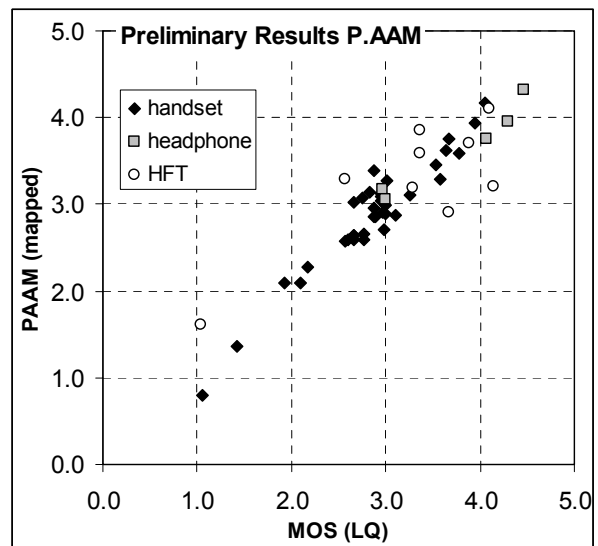


Bild 1: Beispiel für vorläufige P.AAM Ergebnisse

5. Literatur

- [1] "Methods for subjective determination of transmission quality." ITU-T Rec. P.800, 1996.
- [2] "Objective measurement of telephone-band (300-3400 Hz) speech codecs." ITU-T Rec. P.861, 1996.
- [3] "Perceptual evaluation of speech quality (PESQ), an objective method for end-to-end speech quality assessment of narrow-band telephone networks and speech codecs." ITU-T Rec. P.862, 2001.
- [4] Rix, A.W.; Hollier, M.P.; Beerends, J.G.; Hekstra, A.P., "Perceptual Evaluation of Speech Quality (PESQ), the new ITU standard for end-to-end speech quality assessment. Part I – Time-delay compensation." *J. Audio Eng. Soc.*, 50 (10), 755–764, Oct. 2002.
- [5] Beerends, J.G.; Hekstra, A.P.; Rix, A.W.; Hollier, M.P., "Perceptual Evaluation of Speech Quality (PESQ), the new ITU standard for end-to-end speech quality assessment. Part II – Psychoacoustic model." *J. Audio Eng. Soc.*, 50 (10), 765–778, Oct. 2002.
- [6] Rix, A. W.; Reynolds, R.; Hollier, M. P., "Perceptual measurement of end-to-end speech quality over audio and packet-based networks". *106th Audio Engineering Society Convention*, pre-print no. 4873, May 1999.
- [7] Rix, A. W.; Hollier, M. P., "The perceptual analysis measurement system for robust end-to-end speech quality assessment." *IEEE ICASSP*, 3, 1515-1518, June 2000.
- [8] Berger, J., "From end-to-end to mouth-to-ear, speech quality estimation in VoIP scenarios." (in German), *Fortschritte in der Akustik – DAGA 2001, Hamburg / Harburg*.
- [9] Berger, J., "Results of objective speech quality assessment including receiving terminals using the advanced TOSQA2001." *ITU-T SG12*, cont. COM12-20, Dec. 2000.
- [10] Gierlich, H.W.; Kettler, F.; Berger, J.; Klaus, H.; Kliche, I.; Scheerbarth, Th., "Report of 1st ETSI VoIP Speech Quality Test Event.", *ETSI EP TIPHON #22*, 21.–28.03.2001, Bethesda, USA.
- [11] "Anonymized Report of the 2nd ETSI TIPHON VoIP Speech Quality Test Event.", *ETSI*, April 2002, Sophia Antipolis, France.
- [12] Berger, J.; Beerends, J. G.; Rix, A. W., "Preliminary Results for the P.AAM benchmark models." *ITU-T SG12*, cont. COM12-D109, Jan. 2003.