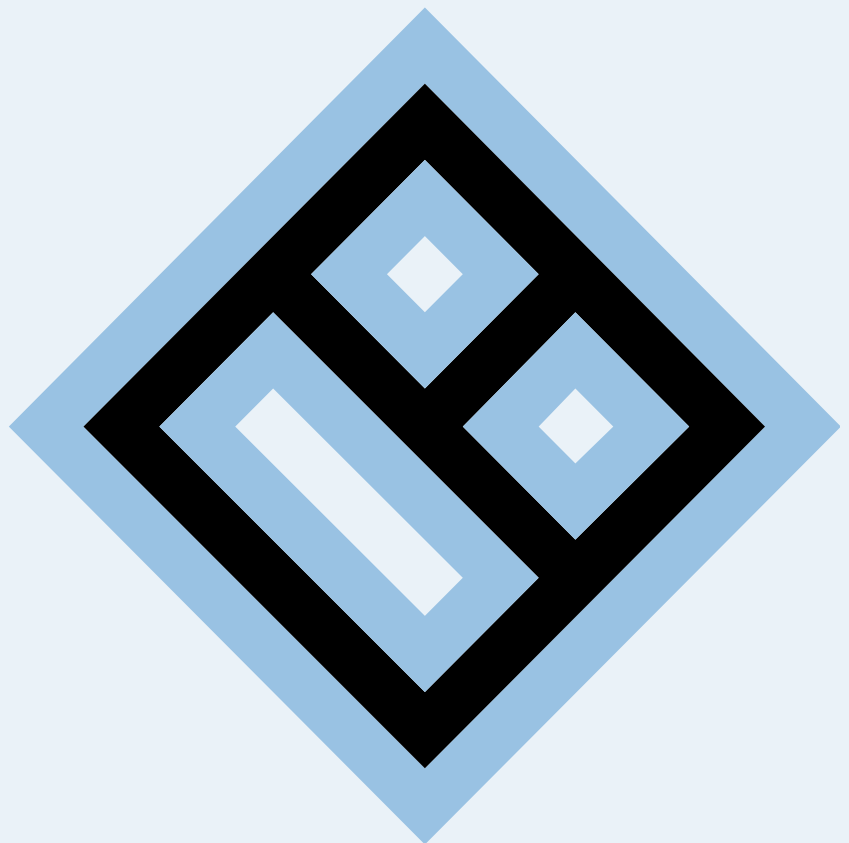


**NEXTRAGEN**  
NEXT GENERATION TESTING

Einsatz von PESQ und  
POLQA in VoIP-Umgebungen:  
ein Vergleich



Kein Teil dieser Broschüre darf in irgendeiner Form (Druck, Fotokopie, Mikrofilm oder in einem anderen Verfahren) ohne unsere vorherige schriftliche Genehmigung reproduziert oder unter Verwendung elektronischer Systeme verarbeitet, vervielfältigt oder verbreitet werden. Wir weisen darauf hin, dass die im Buch verwendeten Bezeichnungen und Markennamen der jeweiligen Firmen im Allgemeinen warenzeichen, marken- oder patentrechtlichem Schutz unterliegen.

Copyright: 2011 Nextragen GmbH  
Stand: 12/2011

Herausgeber:  
Nextragen GmbH  
Lise-Meitner-Str.2  
24941 Flensburg

# Einsatz von PESQ und POLQA in VoIP-Umgebungen: ein Vergleich

## *Management Summary*

Das Ziel dieses Papiers ist es, einen Vergleich des PESQ-Algorithmus und seines Nachfolgers, POLQA, beim Einsatz in VoIP-Umgebungen zu skizzieren. Zunächst werden die beiden QoS-Messverfahren kurz beschrieben und deren Funktionalität in einer Vergleichsstudie untersucht. Anschließend werden die Stärken und Schwächen der beiden QoS-Messmethoden und ihre praktischen Auswirkungen beschrieben.

## Die korrekte Messung der Sprachqualität bei VoIP

Die Implementierung von Breitband-Sprach-Codern (in den Frequenzbereichen von 50 Hz bis 7000 Hz) trägt dazu bei, die Sprachqualität bei VoIP (Voice over IP) erheblich zu verbessern. Um eine hohe Güte der Sprachübermittlung zu erhalten, muss der im Netzwerk zu erreichende Quality of Service (QoS) jedoch überprüft werden. Aus diesem Grund wurden für die Nutzung der Breitband-Codern neue Messmethoden entwickelt. Die QoS Messmethoden unterteilen sich im Wesentlichen in folgende zwei Kategorien:

- Signal-basierte und
- Parameter-basierte Verfahren.

Zur ersten Kategorie gehören der PESQ Algorithmus (Perceptual Evaluation of Speech Quality) und dessen Nachfolger, POLQA (Perceptual Objective Listening Quality Analysis). Die auf diesen Algorithmen aufbauenden Messmethoden bieten den Vorteil, dass diese objektive Messergebnisse und eine sehr genaue Analyse der Sprachqualität liefern. Im Gegensatz dazu sind die Parameter-basierten Messverfahren (beispielsweise das E-Modell oder EIP-Modell) einfach zu bedienen und daher kostengünstig. Die meisten Parameter-basierten Verfahren sind jedoch relativ unzuverlässig und die ermittelten Ergebnisse können erheblich von den tatsächlichen QoS-Werten abweichen.

Das Unternehmen Nextragen GmbH hat den PESQ-Algorithmus bereits seit mehreren Jahren in sein Simulationswerkzeug TraceSim implementiert. Eine erweiterte Version von TraceSim unterstützt jetzt auch Breitband-Codern und die, für die Messungen notwendige Referenz-Datei mit einer 16 kHz Sampling-Frequenz. Außerdem unterstützt TraceSim 2.1 einen PESQ- Nachfolger. Dieser neue Algorithmus heißt „Perceptual Objective Listening Quality Analysis“ (POLQA). Die neue Messmethode ist für Netzwerktests auf einer Ende-zu-Ende-Basis für Super-Breitband Codern (50 bis 14.000 Hz) konzipiert. Hierfür wird bei der Messung eine Referenz-Datei mit Samplingrate von 48 kHz genutzt. Das neue Mess- und Prüfverfahren eröffnet völlig neue Möglichkeiten bei der Durchführung von QoS-Messungen in VoIP-Umgebungen.

## Der PESQ-Algorithmus

Mit dem Aufkommen von Voice over IP (VoIP) wurde die Messung der Sprachqualität eine noch wichtigere, aber auch schwierigere Aufgabe. Anders als in den leitungsgebundenen Telefonnetzen und in den klassischen Mobilfunknetzen schwanken bei Voice over IP die Verzögerungen der Sprachdatenpakete.

Je nachdem an welchem Messpunkt des Endgeräts die Sprachsignale abgegriffen werden (beispielsweise an der Schnittstelle zum IP-Netz oder am Telefonhörer), wirken sich diese Verzögerungsschwankungen (Jitter) mehr oder weniger stark auf das aufgezeichnete Signal aus. Dadurch wird bei Messverfahren mit Eingangs- und Ausgangssignal die Bestimmung der Verzögerung zwischen dem eingespeisten Testsignal und dem Ausgangssignal schwieriger und damit die Analyse der Veränderungen des aufgezeichneten Sprachsignals in Bezug auf das Testsignal.

Die ITU-T hat deshalb im Jahr 2001 in der Empfehlung ITU-T P.862 das PESQ (Perceptual Evaluation of Speech Quality) Verfahren verabschiedet. PESQ ist eine Methode zur objektiven Qualitätsbewertung von Sprache die per Voice over IP über ein Netzwerk übermittelt wird. Dabei werden der klassische Frequenzbereich der Telefonie von 300 bis 3.400 Hz (Schmalband) bzw. die erweiterten Übermittlungsbereiche von 50 bis 7.000 Hz (Breitband) überprüft. Die wichtigsten Eigenschaften des PESQ-Algorithmus sind:

- arbeitet nach dem Standard ITU-T P.862.
- bewirkt eine objektive Bewertung von Telefonverbindungen.
- basiert auf den psychoakustischen Eigenschaften des menschlichen Gehörs.
- ist für Ende-zu-Ende-Messungen ("Full Reference") von VoIP-Verbindungen geeignet.
- ist ein anerkannter Standard zur Messung der Sprachqualität in VoIP-Netzen.

Der PESQ-Algorithmus berechnet die Sprachqualität auf einer Skala von "- 0,5" bis "4,5". Werte in der Region "- 0,5" weisen auf eine sehr schlechte Sprachqualität hin, während Werte im Bereich von "4,5" eine exzellente Sprachqualität darstellen. Die PESQ-Skala lässt sich als MOS (Mean Opinion Score gemäß ITU-T Rec. P.862.1 und P.862.2.) in den Wertebereichen von "1" bis "5" ausdrücken. Die PESQ/MOS-Mapping-Funktion für Schmalband-Codecs lautet:

- $$LQO = 0,999 + \frac{4,999 \text{ bis } 0,999}{1 + e^{-1,4945 \text{ PESQ} + 4,6606}}$$

Für Breitband-Codecs lautet die Gleichung:

- $$LQO = 0,999 + \frac{4,999 \text{ bis } 0,999}{1 + e^{-1,3669 \text{ PESQ} + 3,8224}}$$

## Der POLQA-Algorithmus

Das Messverfahren Perceptual Objective Listening Quality Analysis (POLQA) ist der Nachfolger des PESQ-Algorithmus. Das neue Messverfahren basiert auf dem Standard ITU-T P.863. Der neue Algorithmus wurde zur Qualitätsprüfung von Sprachver-

bindungen entwickelt und eignet sich sowohl für die schmalbandige (300 - 3.400 Hz) als auch für die breitbandige (50 - 14.000 Hz) Untersuchung der übermittelten Sprachsignale. Der POLQA- Algorithmus besitzt zwei Modi:

- Schmalband (Referenzsignal: Abtastrate von 8 kHz) und
- Superbreitband (Referenzsignal: Abtastrate von 48 kHz).

Der Superbreitband-Modus der POLQA-Bewertungsmethode ermöglicht die Bewertung beider Signalklassen auf ein und derselben Skala. Die verminderten Signale, die sich aus der Verwendung eines Codecs mit einer Abtastrate von normalerweise < 48 kHz ergeben, werden intern neu berechnet, um einen Vergleich der beiden Eingangssignale zu ermöglichen. Das heißt, um eine Messung in Übereinstimmung mit dem POLQA Superbreitband-Modus durchführen zu können, muss das Testsignal durch eine Neuberechnung aus dem 48-kHz-Referenzsignal gewonnen werden. Das endgültige Ergebnis der POLQA-Messmethode wird als MOS-LQO-Wert ausgedrückt.

Durch die unterstützten Bandbreiten (Schmalband, Breitband, Superbreitband) der POLQA-Messmethode und der ausschließlichen Verwendung der MOS-Skala (1,0 bis 4,5 für den Schmalband-Modus und 1,0 bis 4,75 für den Superbreitband-Modus) ergeben sich MOS-LQO Werte (gemäß der Norm P.863), die gegenüber den MOS-LQO Werten (gemäß Standard P.862) um ca. 0,5 MOS ausgeglichen werden müssen. Aus diesem Grund lassen sich die beiden Algorithmen auch schwer miteinander vergleichen. Deshalb wird vom POLQA-Algorithmus eine Neuberechnung der MOS-LQO Werte vorgenommen, bevor der endgültige Messwert ausgegeben wird.

Die Neuberechnung erfolgt im Schmalbandbereich nach folgender Formel:

$$a) \text{ MOS LQO} = 0,79 + 0,0036 \cdot \text{MOS P.863} + 0,2117 \cdot \text{MOS P.863} - 0,0065 \cdot \text{MOS P.863}$$

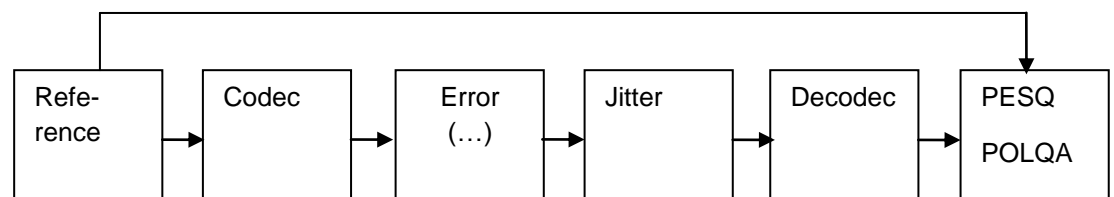
und im Breitbandbereich lautet die Formel:

$$b) \text{ MOS LQO} = 0,276 + 0,7203 \cdot \text{MOS P.863} + 0,00756 \cdot \text{MOS P.863} - 0,01141 \cdot \text{MOS P.863}$$

## Vergleich beider Algorithmen

In der Praxis benötigt man für eine PESQ- oder POLQA-Messung eines 8 Sekunden langen Referenzsignals die Übertragungszeit plus die Verarbeitungszeiten, was sich auf 1 Minute ausdehnen könnte. Um die Messzeiten zu reduzieren, bedient man sich eines Software-Tools, das die Berechnungen numerisch durchführt.

Die folgende Skizze zeigt das Blockschaltbild dieses Werkzeugs.



Nachfolgend wird der grundsätzliche Programmablauf dargestellt, mit dem, in realen Systemumgebungen, die jeweiligen Messverfahren miteinander verglichen werden:

- zuerst wird eine Referenzdatei (analoges Sprachsignal) geladen.

- dieses Signal wird in Einklang mit dem verwendeten Codec (beispielsweise G.711, GSM, etc.) codiert.
- die kodierten Digitaldaten werden entsprechend der gewählten Größe des Sprach-Samples segmentiert und in RTP (Real-Time Transport Protocol)-Paketen verpackt.
- der Baustein "Error" generiert das gewünschte Maß an Beeinträchtigungen (Fehlern) im Netzwerk.
- die empfangenen Pakete werden zunächst im Jitter-Buffer zwischengespeichert und anschließend so bearbeitet, dass in Übereinstimmung mit der Norm ITU-T G.113 die zeitlichen Funktionen angepasst werden.
- die paketierten Sprach-Samples werden entsprechend dem genutzten Sprach-Codec decodiert.
- schließlich werden die Sprachdaten und die Sprachreferenzdatei geladen und an den PESQ- oder POLQA-Algorithmus weitergeleitet.
- zum Abschluss erfolgt die Berechnung der entsprechenden PESQ- oder POLQA-Messwerte.

## Detailvergleich der beiden Messalgorithmen

- a) **PESQ vs. POLQ beim Einsatz im Schmalbandbereich**  
Die folgenden Parameter werden für ordnungsgemäße Analysen vorausgesetzt:

- nichtdeterministische (Binomialverteilung) Verteilung der Paketverluste im Bereich von 0 bis 20 Prozent.
- nichtdeterministische (Exponentialverteilung) Verteilung der Burst-Größe mit einem Mittelwert von 1.
- 8-kHz-Signal (Schmalband-Modus)
- 48-kHz-Signal (Superbreitband-Modus)
- Schmalband-Codec G.711.1 mit einer Bitrate von 80 Kbit/s (Mode R2a) und Sample-Längen von 10, 20 und 30 ms.
- 31 Messungen für jeden Performance-Wert. Auf diese Weise ist es möglich, ein Konfidenzintervall von weniger als 10 Prozent des geschätzten Durchschnitts (mit einer Fehlerwahrscheinlichkeit von 5 Prozent) zu erreichen
- QoS-Werte auf der MOS-LQO Skala.
- "Silence Insertion"-Methode in Fällen von aufgetretenen Paketverlusten.

Die folgenden Abbildungen 2 bis 4 zeigen die Ergebnisse für das erste Messszenario. Die QoS-Kurven entwickeln sich exponentiell. In Fall von Superbreitband-POLQA fallen die Signalfanken steiler ab und erreichen den niedrigen MOS-Wert von ca. 1. Alle drei Algorithmen bestätigen die Qualität des R2a Modus (ca. 4,4 MOS). Auch ist auffällig, dass der Superbreitband-POLQA Modus sensitiver in Bezug auf die Sample-

Längen reagiert als die anderen Algorithmen. Jedoch liefern alle drei Techniken im Großen und Ganzen vergleichbare Ergebnisse und eignen sich gleichermaßen für die Bewertung QoS im Schmalband-Bereich.

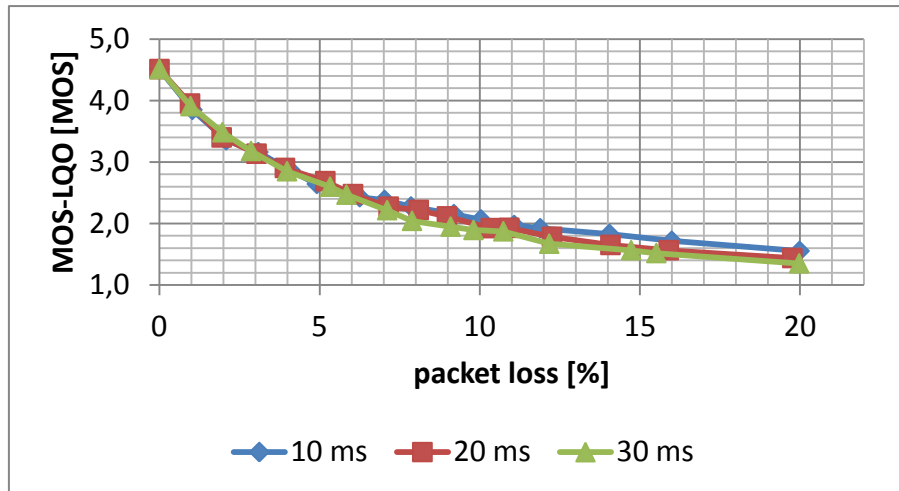


Abbildung 1 MOS-LQO Werte ermittelt von Schmalband-PESQ als Funktion der Paketverluste und Länge der Messprobe für den Codec G.711.1 R2a

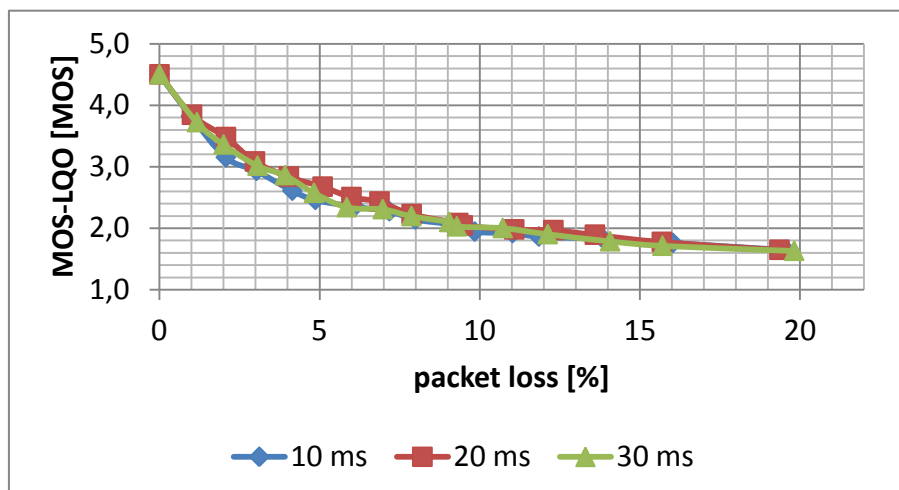


Abbildung 2 MOS-LQO Werte ermittelt von Schmalband-POLQA als Funktion der Paketverluste und Länge der Messprobe für den Codec G.711.1 R2a

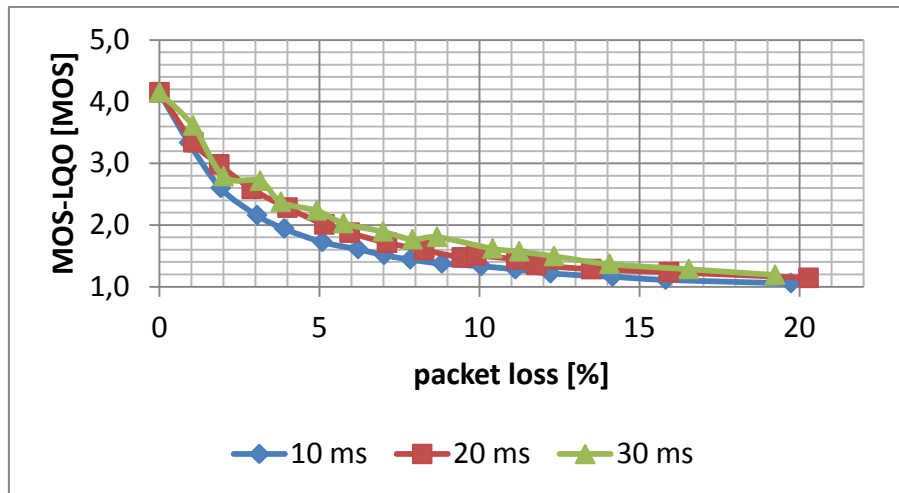


Abbildung 3 MOS-LQO Werte ermittelt von Superbreitband-POLQA als Funktion der Paketverluste und Länge der Messprobe für den Codec G.711.1 R2a

**b) PESQ vs. POLQA beim Einsatz im Breitbandbereich**

Es gelten die gleichen Voraussetzungen wie in Szenario A mit folgenden Ausnahmen:

- 16-kHz-Signal (Breitband-Modus)
- Breitband-Codec G.711.1 mit einer Bitrate von 80 Kbit/s (Mode R2b) und Sample-Längen von 10, 20 und 30 ms.

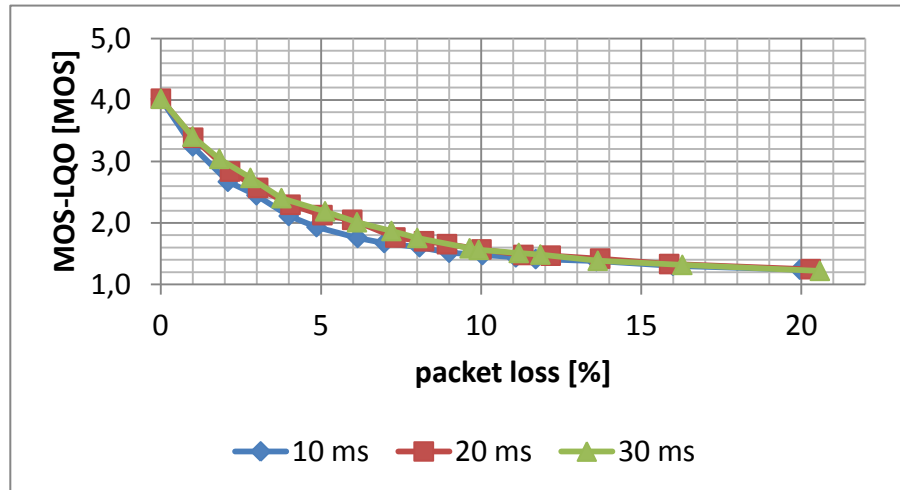


Abbildung 4 MOS-LQO Werte ermittelt von Breitband-PESQ als Funktion der Paketverluste und Länge der Messprobe für den Codec G.711.1 R2a

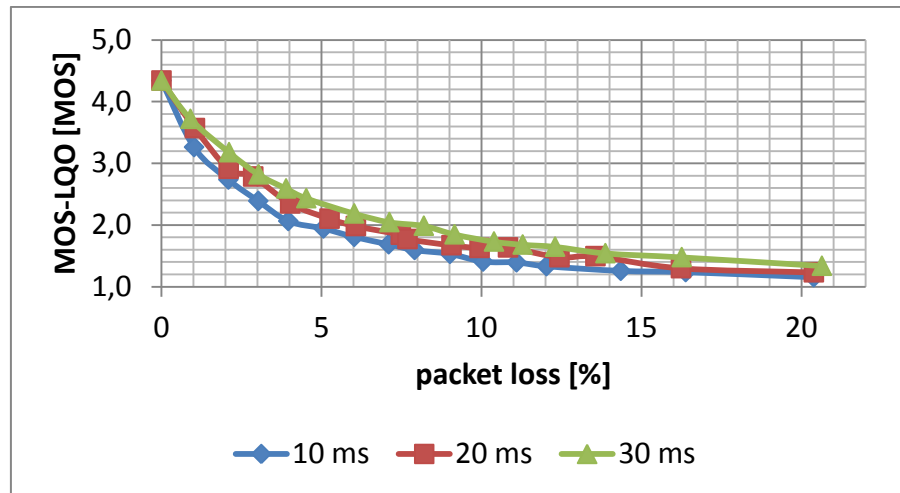


Abbildung 5 MOS-LQO Werte ermittelt von Superbreitband-POLQA als Funktion der Paketverluste und Länge der Messprobe für den Codec G.711.1 R2a

Die QoS-Kurven weisen ebenfalls einen exponentiellen Charakter auf. Beide Algorithmen bestätigen die sehr gute Qualität des R2b-Modus (Breitband-PESQ: MOS ca. 4,0, und Superbreitband-POLQA: MOS ca. 4,4). Der Unterschied von 0,4 MOS ergibt sich aus der Umwandlung der oben dargestellten Formeln (1) und (2), und dem Versuch, die POLQA ermittelten Schmal- und Breitband-Signale in einem einzigen Maßstab darzustellen. Auch hier wird deutlich sichtbar, dass Superbreitband-POLQA sensibler als Breitband-PESQ auf Probenlänge reagiert. Im Großen und Ganzen jedoch ergeben beide Algorithmen vergleichbare Ergebnisse und eignen sich gleichermaßen für die Bewertung QoS in Breitband-Umgebungen.

### Effekte der Pegel

PESQ ist nahezu unempfindlich gegen Niveauunterschiede zwischen dem Referenz- und dem abgeschwächten Empfangssignal. Außerdem werden die Wahrnehmungseffekte der Niveauunterschiede bei PESQ oft unterschätzt. Beide Faktoren wurden bei POLQA korrigiert und die Effekte entsprechen somit den subjektiven Hörtests. Hierzu müssen einige Regeln beachtet werden, damit korrekte Ergebnisse erzielt werden:

- POLQA geht strikt davon aus, dass die Referenzsignale über einen aktiven Sprachpegel von - 26dBov (gemessen nach P.56) verfügen und dieser mit einem nominalen Pegel von 73dB SPL im Bezugspunkt für beide Ohren im Super-Breitband-Modus und 79dB SPL im Bezugspunkt eines Ohres im Schmalband-Modus bereit steht.
- In Super-Breitband-Modus müssen die reduzierten Signale zur Validierung mit POLQA einen aktiven Pegelbereich zwischen - 21dBov bis -46dBov aufweisen. Nur dann arbeitet POLQA zuverlässig. Ein Pegelniveau von -26dBov repräsentiert die ideale Normallautstärke.
- Im Schmalband-Modus sollte das verminderte Signal ebenfalls 26dBov entsprechen.

Es ist zu beachten, dass das häufig verwendete, zur Referenzmessung genutzte "Cool Edit Pro" Werkzeug nicht gemäß P.56 misst und deshalb die daraus resultierenden Messwerte um bis zu 3 dB variieren können.

## Schmalband - Breitband - Super-Breitband

Bei PESQ gibt es nur zwei Betriebsarten:

- Schmalband (300 – 3400 Hz) und
- Breitband (100 - 7000 Hz).

Für POLQA sind ebenfalls zwei Modi definiert. Diese decken jedoch den Schmalband- und den Super-Breitband- (50 -14000 Hz) Bereich ab.

Dadurch entsteht eine Überschneidung zwischen PESQ und POLQA im Schmalbandbereich, aber im Breitband- und Super-Breitband-Bereich gibt es bei PESQ keine direkte Entsprechung.

Aufgrund der genutzten Algorithmen liefern PESQ und POLQA im Schmalband zwar in der Regel ähnliche Ergebnisse, ein direkter Vergleich sollte aber vermieden werden. Der Breitbandmodus von PESQ wird vollständig vom POLQA Super-Breitband-Modus ersetzt. In der Konsequenz heißt dies für die korrekte Anwendung von POLQA für Breitband-Szenarien: Es müssen immer die Super-Breitband-Referenz-Dateien für die Tests verwendet werden. Die dadurch erreichbaren Ergebnisse werden im Gegensatz zum Breitband-Modus von PESQ auf einem anderen Bereich der MOS-Skala abgebildet. Dies hat seine Ursache in den unterschiedlichen Voraussetzungen der individuellen Schmalband-, Breitband- und Super-Breitband-Modi und der daraus resultierenden unterschiedlichen Abbildung auf der MOS-Skala. Bei der Entwicklung des POLQA-Algorithmus wurde darauf geachtet, dass eine Messung mit POLQA unter typischen Breitband-Bedingungen einen ähnlichen Wertebereich (Maximalwert 4,5) ergibt, wie Messungen von PESQ im Breitband-Bereich.

Zum direkten Vergleich zwischen PESQ- und POLQA-Ergebnisse muss eine Transformationsskala angewendet werden. Dabei werden die POLQA-Werte an den anderen Kontext (Breitband vs. Super Breitband) angepasst.

## MOS-Wertebereiche

Die Tabelle 1 zeigt die Grenzen der MOS Skala bei PESQ (P.862.1/2) und bei POLQA in Abhängigkeit von der jeweiligen Betriebsart und des eingesetzten Algorithmus. Dabei ist zu beachten, dass es keine Überschneidungen für den Breitband und den Super Breitband-Modus gibt. Im Super Breitband-Modus wird POLQA bei einem ausgezeichneten Breitbandsignal im Vergleich zu einem Super Breitband-Referenzsignal einen Messwert von ca. 4,5 liefern. In gleicher Weise wird ein perfektes Schmalbandsignal im Vergleich zu einem entsprechenden Super Breitband-Referenzsignal einen MOS-Wert von ca. 4,0 ermitteln.

Modus	P.862.1/2	P.862.1/2	POLQA	POLQA
	MOSmin	MOSmax	MOSmin	MOSmax
Schmalband	1	4,5	1	4,5
Breitband	1	4,5	-	-
Super-Breitband	-	-	1	4,75

Tabelle 1: Grenzen der jeweiligen MOS-Skalen

PESQ ist sehr unempfindlich gegenüber linearen Verzerrungen des Frequenzgangs. Diese hatten jedoch nur Auswirkungen auf die Messergebnisse in Extrembereichen. Mit POLQA wurde dieser Problembereich behoben und die Verzerrungen entsprechen jetzt der menschlichen Wahrnehmung.

PESQ ist sehr tolerant gegenüber falschen Aufnahmepegeln. Dagegen sorgt POLQA für die genaue Einhaltung der Pegel. PESQ reagiert auch sehr empfindlich auf Taktunterschiede zwischen D/A- und A/D-Wandler im Übertragungspfad. Dies führt in der Praxis oft zu schlechteren Messergebnissen. POLQA dagegen ist in der Lage, solche Probleme zu korrigieren bzw. zu kompensieren.

### Zusammenfassung und Ausblick

Die neue ITU-T Recommendation P.863 „POLQA“ stellt ein Technologie-Update für den Standard P.862 „PESQ“ dar und verbessert die Qualität einer Ende-zu-Ende Messung im VoIP-Bereich signifikant. Da POLQA die gleichen Skalen und Bewertungsmechanismen wie PESQ benutzt, sind die jeweiligen Messergebnisse miteinander vergleichbar. Somit lassen sich ältere PESQ-Messreihen auf Basis von neueren POLQA-Messreihen fortsetzen.

Die Messprodukte der Nextragen GmbH unterstützen beide Messverfahren und können deshalb in jeder VoIP-Umgebung eingesetzt werden.

### Firmenprofil Nextragen

Die Nextragen GmbH ist auf die Entwicklung von VoIP/Video-Produkten und Software-Lösungen rund um die Themen „Monitoring, Analysing und Testing“ ausgerichtet. Die Sicherstellung der End2End Dienstqualität (QoS, QoE) für Next Generation Networks und Triple Play Dienste steht im Fokus des im Jahr 2009 gegründeten Unternehmens mit Sitz an der Flensburger Förde im Norden Deutschlands. Kunden aus dem Carrier-, Telekommunikations- und Enterprise-Segment setzen die Nextragen-Lösungen ein, um die Qualität und Zuverlässigkeit von VoIP- und Video-Anwendungen zu monitoren, zu analysieren und zu testen. Produkte, Lösungen und Dienstleistungen der Nextragen GmbH sind 100% „made in Germany“ und werden weltweit über zertifizierte Partner vertrieben.

Weitere Informationen erhalten Sie auf der Firmenwebsite unter [www.nextragen.de](http://www.nextragen.de).



Nextragen GmbH  
Lise-Meitner-Str.2  
24941 Flensburg

Telefon: +49 461 9041-4440  
Fax: +49 461 9041-4449

[www.nextragen.de](http://www.nextragen.de)  
[info@nextragen.de](mailto:info@nextragen.de)

Änderungen und Irrtümer vorbehalten