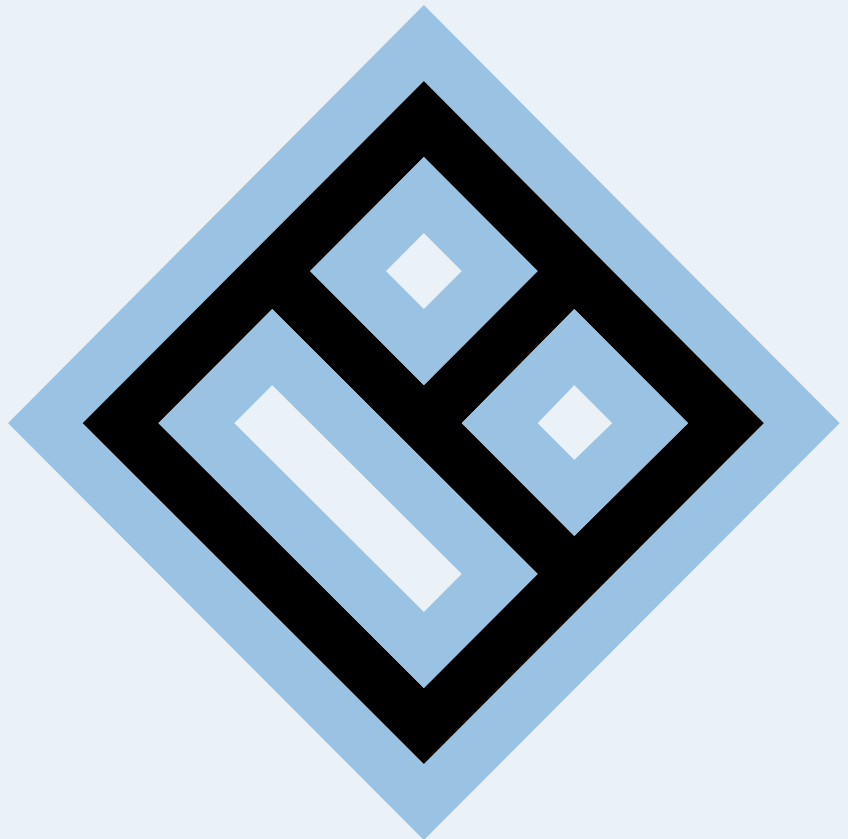


NEXTRAGEN

NEXT GENERATION TESTING

So wird die Sprachqualität in VoIP-Gateways richtig gemessen



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: 2012 Nextragen GmbH
Stand: 02/2012

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

Management Summary

Die heutige Telekommunikationsinfrastruktur der Unternehmen wandelt sich. VoIP wird vielfach flächendeckend eingesetzt. Neu geplante Gebäude werden mit modernster VoIP-Technik ausgestattet und alte TK-Anlagen werden durch VoIP ersetzt. Dies bietet viel Komfort in der Bedienung und erweitert die technischen Möglichkeiten des Unified Communications. Da aufgrund der bestehenden Infrastrukturen in den öffentlichen Netzen (PSTN bzw. TDM-Netze) vielfach noch keine durchgängige (Ende-zu-Ende) VoIP-Techniken zu realisieren sind, müssen an den Netzübergängen spezielle Gateways installiert werden. Die Gateways können einen maßgeblichen Einfluss auf die Sprachqualität haben und die Überprüfung der VoIP-Ströme erschweren. Ohne das notwendige Wissen und die richtigen Messwerkzeuge täuschen die Messwerte ein falsches Messergebnis vor. Im Folgenden Nextragen Whitepaper wird dargestellt, welche Probleme Gateways für die Messtechnik darstellen und wie diese mit den richtigen Messmethoden überwunden werden können.

Jedes Gerät, das für die Sprachübermittlung in Telekommunikationsnetzen (PSTN, Voice over IP (VoIP) oder Mobilfunk) verwendet wird, muss eine gute Sprachqualität bieten. Da es sich bei der Sprachqualität um ein mehrdimensionales und komplexes Problem handelt, muss der Techniker wissen, wie diese bestimmt wird. Betrachten wir zum Beispiel ein typisches Gespräch zwischen zwei Personen über das Telefon. Bei jeder der am Gespräch beteiligten Personen handelt es sich sowohl um einen Sprecher als auch einen Zuhörer. Die Konversation über das Telefon zwischen zwei Menschen lässt sich auf Szenarien reduzieren.

Szenario 1

Das Szenario 1 wird als Single-Talk-Szenario bezeichnet. Das Gespräch wird nur von Person A geführt. Diese Person redet, während der andere Teilnehmer (Person B) nur zuhört. In diesem Szenario gibt es zwei Perspektiven in Bezug auf die über die Telefonverbindung erzielte Sprachqualität - die Perspektive des Sprechers und die Perspektive des Zuhörers. Der Sprecher A hört beim Sprechen sein eigenes Echo (falls vorhanden) und im gewissen Maß auch Umweltstörungen. Aus diesen Informationen beurteilt der Sprecher A die Sprachqualität. Der Zuhörer B befindet sich im selben Szenario im Hörmodus und empfängt nur die übertragenen Sprachinformationen von Person A. Das Echo, welches Person A vernimmt, bekommt Person B nicht mit. Auch die Umweltstörungen werden von Person B nicht wahrgenommen, dass diese von der übermittelten Sprache überlagert werden. Somit orientiert sich die Beurteilung der Sprachqualität von Person B an der von Person A empfangene Sprachinformationen.

Beide Personen beurteilen in diesem Szenario die Sprachqualität aus einer anderen Perspektive. Beide Beurteilungen sind gleichermaßen gültig. Daher müssen beide abgefragt werden, um ein endgültiges Urteil über die für das Telefonat erzielte Sprachqualität zu erhalten.

Szenario 2

Szenario 2 wird als Double-Talk-Szenario bezeichnet. In dieser Situation sprechen beide Parteien zur gleichen Zeit und beurteilen individuell die erzielte Sprachqualität. Die Urteile der beiden Personen über die erzielte Sprachqualität setzen sich aus folgenden Komponenten zusammen:

- Die Qualität der jeweiligen Sprachübertragung zur Person am anderen Ende der Telefonverbindung und

- die Auswirkungen des eigenen Echos und der Umweltgeräusche auf die Sprachqualität.

Die Beurteilung der Sprachqualität ist in diesem Szenario erheblich komplizierter als in Szenario 1. Beide Personen sprechen und hören und versuchen gleichzeitig die Sprachqualität zu beurteilen. Diese Beurteilung ist schwierig, da die Aufmerksamkeit der handelnden Personen durch ihre Tätigkeiten (Sprechen und gleichzeitigem Hören) überlagert wird.

Definition der Sprachqualität

Was ist Sprachqualität? Ist das Qualitätsurteil eines Single- oder Double-Talk-Szenarios ausschlaggebend oder werden beide Szenarien zur Urteilsbildung herangezogen? Wird die Qualität aus der Perspektive des Sprechers oder Hörers gebildet oder werden beide Szenarien zur Beurteilung herangezogen? Welche Facetten der Sprachqualität bilden die Grundlage der Beurteilung eines Telefongesprächs zwischen zwei Menschen?

Die Beurteilung der Sprachqualität bei der Übermittlung von Telefongesprächen gehört zu den komplexen und mehrdimensionalen Funktionen, da es sich in diesem Fall um eine Kombination von Sprachübertragungstechnologie und komplexen menschlichen Gesprächsfunktionen handelt.

Zur technischen Sprachbeurteilung müssen daher Definitionen der Sprachqualität festgelegt werden, die die entsprechenden Metriken der Sprachqualität und deren Wirkungen beschreibt. Die Sprachqualität ist somit das Ergebnis einer allgemein akzeptierten Beurteilungsskala auf der ein mehrdimensionales Sprachsignal abgebildet wird. Diese beschreibt den subjektiven Wahrnehmungseindruck eines Menschen auf die von ihm wahrgenommenen Sprachinformationen. Damit gehört die Beurteilung der Sprachqualität zu den psycho-akustischen Urteilen. Somit wird die Sprachqualität nach subjektiven Kriterien ermittelt, indem Sprechproben den Probanden vorgespielt, die einzelnen Bewertungen gewichtet und daraus die statistischen Ergebnisse ermittelt werden. Um repräsentative Zahlenwerte zu erhalten, muss ein Publikum von mindestens 100 Menschen befragt werden. Diesem Publikum wird eine Sprachprobe vorgespielt. Jeder Proband vergibt nach dem Hören der Sprachprobe der verminderten Sprachqualität eine Note. Aus den vergebenen 100 Noten wird anschließend der Durchschnittswert errechnet.

Aus praktischen Erwägungen nutzt man heute international festgelegte Messstandards. Diese wurden von den internationalen Telekommunikationsgremien (ITU-T, ETSI, etc) verabschiedet. Typischerweise wird dabei die so genannte „Toll-Quality“ - die Sprachqualität von PSTN-Systemen referenziert. Diese subjektiven Beurteilungsmethoden wurden in dem Sprachqualitätsstandard ITU-T P.800 festgeschrieben.

Der Begriff der Sprachqualität wird in Messungen der Ein-Wege-Sprachqualität und der Zwei-Wege-Sprachqualität unterschieden. Beide Maßnahmen bieten unterschiedliche Informationen zur Beurteilung der Sprachqualität.

Zur Beurteilung eines Gesprächs oder einer Zwei-Wege-Sprachqualität wird die Sprachqualität eines Gespräches als Gesamtes bewertet und entspricht der Bewertung der Sprachqualität in einem Double-Talk-Szenario.

Es beinhaltet die Auswirkungen aller Komponenten des Kommunikationsnetzes. Für die Übermittlung von Sprache über ein digitales Übermittlungsnetz wird ein so genannter Codec (Codierer/Decodierer) genutzt. Der verwendete Sprach-Codec hat

direkte Auswirkungen auf die Signalschwankungen am Eingang des Sprach-Codex. Die Hintergrundgeräusche, die Auswirkungen von Paketverlusten, die Verzögerungen, den Jitter und das Echo sind Auswirkungen, welche während der Übertragung entstehen.

Der wesentliche Vorteil einer umfassenden Sprachqualitätsmessung besteht darin, dass man ein Beurteilungsmaß für die gesamte Konversation und somit ein Bewertungsmaß des gesamten tatsächlich verwendeten Telekommunikationssystems erhält. Dieser Lösungsansatz hat jedoch zwei wesentliche Nachteile:

1. In der Praxis kann nur sehr schwierig festgestellt werden, welche Komponenten im Telekommunikationssystem das schwächste Glied in der Gesamtkommunikation ist. Die alleinige Messung der Sprachqualität eines gesamten Übertragungskanals ist nicht ausreichend, um Auskunft über die Auswirkungen der einzelnen Komponenten des Telekommunikationssystems auf die Sprachqualität zu erhalten. Daher eignet sich dieses Messverfahren nicht zur Ermittlung von Einzelkomponenten, die die Sprachqualität negativ verändern. Um die Auswirkungen der einzelnen Komponenten eines Telekommunikationssystems auf die Sprachqualität feststellen zu können, muss die Sprachqualität bzw. die durch die betreffende Komponente hervorgerufene Verminderung separat gemessen werden.
2. Es gibt keine objektiven Messwerkzeuge die vorhersagen können, wie das menschliche Gehör die Stimme eines Gesprächspartners beurteilt. Daher erfordert eine allgemein gültige Sprachqualitätsprüfung spezielle Messmethoden und spezielle Messeinrichtungen.

Im Gegensatz dazu beurteilt die Einwege-Sprachqualität die Qualität der Sprache aus der Perspektive des Zuhörers. Per Definition deckt dieses Messverfahren nicht die Auswirkungen des Echokompensators auf Sprachqualität noch die Auswirkungen der Verzögerung auf die Zweibege-Sprachqualität ab. Allerdings werden die Auswirkungen aller anderen relevanten Sprachparameter und aller Komponenten im Netzwerk erfasst.

Die Messung der Sprachqualität auf Basis des Einwege-Konzepts bietet drei wesentliche Vorteile:

1. Die Sprachqualität eines Großteils der Telekommunikationsparameter (mit Ausnahme der Effekte des Echokompensators) lassen sich problemlos ermitteln.
2. Die zur Beurteilung einer Konversation notwendigen Parameter ergeben automatisch die Werte zur Beurteilung der Sprachqualität.
3. Inzwischen stehen international anerkannte objektive Algorithmen (beispielsweise ITU-T P.862) zur Beurteilung der Sprachqualität zur Verfügung. Diese bilden das menschliche Hörvermögen für den Einsatz in Telekommunikationssystemen ab.

Die Verfügbarkeit von zuverlässigen objektiven Messvorschriften reduziert gegenüber den subjektiven Messmethoden mit menschlichen Probanden, die für die Tests notwendige Zeit und trägt somit zur erheblichen Kostenreduzierung bei. Von größerer Bedeutung sind diese Messverfahren bei der Beurteilung von Voice-Gateways, da immer die gleichen objektiven Maßstäbe zur Beurteilung der Sprachqualität herangezogen werden.

Echokompensation und ihre Auswirkungen auf die Qualität der Stimme

Echokompensatoren spielen eine große Rolle bei der Beurteilung der Sprachqualität von Konversationen mit Hilfe von Telekommunikationsgeräten. Der ITU-T Standard ITU-T G.131 dient der Beurteilung der Auswirkungen des Echos auf die Sprachquali-

tät. Die Auswirkungen des Echos auf die Sprachqualität hängen von der Höhe des Echo-Signals und ihrer Verzögerung ab.

Bis heute gibt es keine objektiven Beurteilungswerkzeuge, die die Auswirkungen der Echokompensatoren auf eine Konversation eines menschlichen Zuhörers beschreiben. Aus diesem Grund wird die Echokompensation in folgenden zwei Schritten bewertet:

1. Auf Basis einer objektiven Messung wird die Leistung der Echounterdrückung (gemäß der ITU-T Rec. G.168) ermittelt.
2. Beurteilt die Wirkung von Echokompensatoren auf die Sprachqualität unter Verwendung realer menschliche Hörer oder auf Basis von subjektiven Messmethoden (P.831).

Die ITU-T G.168 Empfehlung spezifiziert die Anforderungen einer netzspezifischen Echounterdrückung und legen eine Reihe von objektiven Messgrößen fest. Entspricht die Echounterdrückung den in G.168 festgeschriebenen Werten, dann garantieren diese ein Mindestmaß an Leistungen. Hierzu gehören niedrige bzw. hohe Hintergrundgeräusche, niedrige bzw. hohe Signalverdoppelungen, variable Übermittlungszeiten, nicht lineare Prozessoreffekte (NLP), Rückflusdämpfung, und die Stabilität der Signale. Die G.168 Spezifikation legt nur die objektive Anforderungen an die Echounterdrückung fest und beseitigt jedoch nicht die Auswirkungen der Echokompensatoren auf die Sprachqualität. Die Auswirkungen auf Sprachqualität werden nicht allein von der Echounterdrückung bestimmt. Auch die Verzögerung, der Jitter und andere Beeinträchtigungen im Netzwerk haben einen hohen Einfluss auf die Leistung des Sprachkommunikation. Dies bedeutet, dass die Echounterdrückung mit den Effekten der Netzkomponenten interagiert. Bis heute gibt es keine objektiven Messmethoden, die ähnlich wie der menschliche Zuhörer, die Effekte der Echokompensatoren auf Sprachqualität bemessen kann. Aus diesem Grund müssen die G.168 Spezifikation und andere objektive Testverfahren durch subjektive Testverfahren ergänzt werden.

Im zweiten Schritt wird die Wirkung der Echokompensatoren auf die Sprachqualität mit Hilfe von subjektiven Methoden beurteilt. Hierzu wird gemäß ITU-T Empfehlung P.831 eine Gruppe von menschlichen Testhörern genutzt. Die ITU-T P.831 beschreibt vier verschiedene Methoden zur Beurteilung der Auswirkungen der Echokompensation auf die Sprachqualität.

Durch die Kombination von objektiven (G.168) und subjektive Messmethoden (P.831) erhält man einen Mechanismus zur Bestimmung der Leistung der Echounterdrückung und deren Auswirkungen auf die Sprachqualität.

Die Auswirkungen auf die Sprachqualität durch Verzögerung werden in der Regel auf Basis von Zweiwege-Sprachqualitätstests überprüft. Die Verzögerung hat entweder nur einen geringfügige bzw. keine Auswirkungen auf die Einwege-Sprachqualität. Dagegen hat die Verzögerung in bidirektionalen (Zweiwege) Gesprächen einen hohen Einfluss auf die Gesprächsqualität. Das Dokument ITU-T G.114 besagt, dass in einer unidirektionalen (Einwege) Übertragung die Verzögerung nicht mehr als 150 ms betragen darf, um vom Nutzer als „noch akzeptabel“ klassifiziert zu werden. Verzögerungen zwischen 150 und 400 ms werden als "akzeptabel mit Einschränkungen" bewertet. Verzögerungen von mehr als 400 ms gelten dagegen als "nicht hinnehmbar." Momentan gibt es jedoch keine international akzeptierten objektiven Messmethoden zur Überprüfung der Auswirkungen von Verzögerungen auf Zweiwege-Gespräch.

Lösungsvarianten zur Messung von Sprachqualität

Wie lässt sich die Sprachqualität eines Telekommunikationsgeräts oder -systems beurteilen? Heute werden folgende zwei Lösungsansätze zur Messung der Sprachqualität genutzt: Die subjektiven und die objektiven Messmethoden.

Subjektive Methoden zur Messung der Sprachqualität

Die Sprachqualität beschreibt, wie gut die Verständlichkeit einer menschlichen Stimme bei Aufzeichnung und Wiedergabe durch eine technische Einrichtung ist. Eine Beurteilung der Sprachqualität ist dabei subjektiv und hängt sowohl von den gegebenen technischen Mitteln, dem Umfeld der Aufnahme, dem Übertragungsweg und dem Umfeld der Wiedergabe ab. Die Bewertung dieser Sprachqualität ist durch Bewertungsmethoden der ITU mit dem Standard P.800 spezifiziert.

Mean Opinion Score (MOS)

Das bekannteste Verfahren zur Bewertung der Sprachqualität ist der »Mean Opinion Score«, kurz MOS. Die ITU hat die Technik Mean Opinion Score im Jahre 1998 als Empfehlung P.800 veröffentlicht. Diese Methode beschreibt die subjektive Wahrnehmung der verschiedenen »Prüflinge«, beispielsweise Codecs mit Hilfe einer festgelegten Skala zur Beurteilung der QoS-Empfindungswerte. Der MOS-Wert ist ein dimensionsloser Wert ähnlich den Schulnoten zwischen eins und fünf. Dabei steht der Wert »1« für eine mangelhafte Sprachqualität, bei der keine Verständigung möglich ist, der Wert »5« hingegen signalisiert eine exzellente Übertragungsqualität, die nicht von dem Original zu unterscheiden ist.

Der MOS-Wert wird subjektiv ermittelt, indem Sprechproben Probanden vorgespielt, die einzelnen Bewertungen gewichtet und daraus die statistischen Ergebnisse ermittelt werden. Die einzelnen Schritte bei der MOS-Technik ergeben sich wie folgt:

- Es wird ein Publikum von mindestens 100 Menschen befragt.
- Danach wird eine Sprachprobe mit dem entsprechenden Sprachcodec kodiert und diese Sprachsequenz dem Publikum vorgespielt.
- Jeder Einzelne vergibt danach für die verminderte Qualität der Sprachprobe eine Note. Diese muss zwischen 1 (schlecht) und 5 (ausgezeichnet) liegen.
- Aus den 100 Noten wird der Durchschnitt berechnet und als MOS-Wert für den Sprach-Codec festgelegt.

Die nachfolgende Tabelle zeigt die gängigsten Codecs und den ermittelten MOS-Wert. Die abgebildeten MOS-Werte entsprechen der besten Qualität, die ein Sprachcodec erhalten kann.

Codec	Mean Opinion Score (MOS)
G.711	4,4
G.729	3,92
G.726	3,85

iLBC	3,8
G.729a	3,7
G.723.1	3,65
G.728	3,61

Tabelle: Codecs und deren dazugehörigen MOS-Werten

MOS bedeutet in seiner Ursprungsform Mean Opinion Score, also durchschnittlicher Meinungswert. Er wird von vielen Verfahren und Algorithmen als Ausgabewert benutzt. Diese Werte lassen sich nicht immer genau vergleichen, da die unterschiedlichen Verfahren differierende Ansätze haben. Deshalb hat die ITU in der Veröffentlichung Rec. P.800.1 beschrieben, dass dem Wort "MOS" noch Zeichen angehängt werden müssen, um eindeutig zu zeigen, welche Art von Verfahren dem berechneten MOS-Wert zu Grunde liegt.

In dieser Empfehlung werden folgende Zeichen erklärt, die an das Wort MOS anzuhängen sind:

LQ – Listen Quality (Gehörte Qualität)

CQ – Conversational Quality (Dialog-Qualität)

S – Subjective

O – Objective

E – Estimated (Geschätzt)

Die nachfolgende Tabelle zeigt die daraus resultierenden Kombinationen:

	Listening-only	Conversational
Subjektive	MOS-LQS	MOS-CQS
Objektive	MOS-LQO	MOS-CQO
Geschätzte	MOS-LQE	MOS-CQE

Tabelle: Verfügbare MOS-Terminologien

MOS-LQS: Dieser Wert wurde in einem Labor erhoben. Das arithmetische Mittel wird aus gesammelten subjektiven Beurteilungen berechnet, welche zwischen 1 und 5 liegen, wie in der Empfehlung Rec. P.800 beschrieben. Ergebnisse basierend auf der Empfehlung Rec. P.830 liefern MOS-LQS.

MOS-LQO: Dieser Wert soll einem Test entsprechen, der objektiv ist und einem gehörten Gespräch entspricht. Das Verfahren in der Empfehlung P.862 entspricht einem solchen Test.

MOS-LQE: Dieser Wert soll von einem Netzwerkplanungsmodell berechnet werden. Dieses Modell muss das Ziel haben, die Qualität einer zu hörenden Applikation wiederzugeben.

MOS-CQS: Dieser Wert wurde in einem Labor erhoben. Das arithmetische Mittel wird aus gesammelten subjektiven Beurteilungen berechnet, welche zwischen 1 und 5 liegen, wie in der Empfehlung Rec. P.800 beschrieben. Subjektive Konversationstests wie in der Empfehlung P.800 liefern MOS-CQS.

MOS-CQO: Dieser Wert soll verwendet werden, wenn das berechnende Modell das Ziel hat, eine objektive Qualitätsbeurteilung eines Gesprächs zu ermitteln. Ein Beispiel dafür ist die Empfehlung Rec. P.562.

MOS-CQE: Bei einem Netzwerkplanungsmodell, welches der Qualität für Applikationen zur Konversation berechnet, wird MOS-CQE benutzt. Als Beispiel ist hier das Modell in der Empfehlung Rec. G.107 nach der Umrechnung zu nennen.

Der Zusammenhang von MOS-LQS, MOS-LQO und MOS-LQE ergibt sich wie folgt:

- Wird das zu testende Signal von einem Menschen subjektiv beurteilt, entspricht dies dem MOS-LQS.
- Wenn das gleiche Signal einem objektiven Algorithmus zur Verfügung gestellt wird, berechnet dieser einen MOS-LQO.
- Aus dem System, über das das Signal gesendet wird, können ermittelte Parameter an einen Algorithmus gegeben werden. Dieser hat eine Datenbank von differierenden Faktoren und kann daraus einen MOS-LQE liefern.

E-Modell

Das E-Modell ist in der ITU-Empfehlung G.107 spezifiziert. Es beschreibt ein Berechnungsmodell zur Planung und Bewertung der Übertragungsqualität von Kommunikationsnetzen. Anhand dieses Berechnungsmodells wird die dem Nutzer in einer Verbindung zur Verfügung stehende Sprachqualität ermittelt. Das Ergebnis ist eine objektive Bewertung der Übertragungsqualität unter Berücksichtigung aller, die Übertragungsqualität beeinflussender Faktoren. Die nachfolgende Tabelle listet die in das E-Modell einfließenden Parameter auf.

Parameter	Abkürzung	Einheit	Standardwert	Gültiger Bereich
Send Loudness Rating	SLR	dB	+8	0 bis +18
Receive Loudness Rating	RLR	dB	+2	-5 bis +14
Sidetone Masking Rating	STMR	dB	15	10 bis 20
Listener Sidetone Rating	LSTR	dB	18	13 bis 23
D-Value of Telephone, Send Side	Ds	-	- 3	-3 bis +3
D-Value of Tele-	Dr	-	- 3	-3 bis +3

phone, Receive Side				
Talker Echo Loudness Rating	TELRL	dB	65	5 bis 65
Weighted Echo Path Loss	WEPL	dB	110	5 bis 110
Mean one-way Delay of the Echo Path	T	ms	0	0 . . . 500
Round Trip Delay in a 4-wire Loop	Tr	ms	0	0 bis 1000
Absolute Delay in echo-free Connections	Ta	ms	0	0 bis 500
Number of Quantization Distortion Units	qdu	-	1	1 bis 14
Equipment Impairment Factor	le	-	0	0 bis 40
Packetloss Robustness Factor	Bpl	-	1	1 bis 40
Random Packetloss Probability	Ppl	%	0	0 bis 20
Burst Ratio	BurstR	-	1	1 bis 2
Circuit Noise referred to 0 dBm-point	Nc	dBm0p	-70	-80 bis -40
Noise Floor at the Receive Side	Nfor	dBmp	-64	-
Room Noise at the Send Side	Ps	dB(A)	35	35 bis 85
Room Noise at the Receive Side	Pr	dB(A)	35	35 bis 85
Advantage Factor	A	-	0	0 bis 20

Tabelle: Parameter des E-Modells

Das E-Modell verwendet für die Bestimmung der Sprachqualität ein passives Modell. Das Messsystem berechnet aus einem übermittelten VoIP-Strom die für das E-Modell notwendigen Parameter. Nach der Übergabe der Parameter an das E-Modell gibt das Messsystem einen Übertragungsfaktor (R-Faktor) aus. Aus diesen Werten wird eine Vorhersage der Sprachqualität im Bereich 0 bis 100 getroffen, die auf der MOS-Skala abbildbar ist.

R-Faktor	Qualität	MOS
100	ausgezeichnet: Es ist keine Anstrengung nötig, um die Sprache zu verstehen	5
80	gut: Durch aufmerksames Hören kann die Sprache ohne Anstrengung wahrgenommen werden	4.0
60	ordentlich: Die Sprache kann mit leichter Anstrengung wahrgenommen werden	3.1
50	mäßig: Es bedarf großer Konzentration und Anstrengung, um die übermittelte Sprache zu verstehen	2,6
bis 0	mangelhaft: Trotz großer Anstrengung kann man sich nicht verständigen	bis 1

Tabelle: Vergleich R-Faktor und MOS

Das E-Model wurde ursprünglich als Planungsparameter für Telefonnetzbetreiber entwickelt. Inzwischen hat es sich als Quasi-Standard zur objektiven Beurteilung der Sprachqualität (im Gegensatz zur subjektiven Meßmethode des MOS) durchgesetzt. Da sich der R-Faktor direkt auf den aus den Tests generierten Messwerten abbilden lässt, entspricht dieser Wert den realen Verkehrsparametern. Trotzdem ist eine Korrelation mit den MOS-Werten möglich. Der beste zu erreichende theoretische R-Faktor beträgt 100. Dieser Wert berücksichtigt jedoch nicht die genutzten Codecs. Nutzt man beispielsweise einen typischen G.711-Codec in einer Referenzumgebung, kann ein maximaler R-Faktor von ungefähr 93,2 erreicht werden. Die folgenden Ursachen tragen zu einer Verschlechterung des R-Faktors bei:

- Codec-Typ: Codecs mit höheren Kompressionsraten weisen normalerweise einen schlechteren R-Faktor auf
- Verfügbare Bandbreite: Einschränkung der Übertragungsbandbreite werden durch das gesamte Übermittlungssystem im Übertragungspfad bestimmt
- Verzögerungen und Jitter: Entstehen im Netzwerk und den Endgeräten und sind bei mobilen WLAN-Telefonen, aufgrund der mangelnden Bandbreite im Funknetz, besonders hoch
- Paketverluste: entstehen durch überlastete Netzwerke und defekten Komponenten

Die Vorteile des E-Modells liegen in seiner Einfachheit. Es muss lediglich der Datenstrom aufgezeichnet werden und die Parameter ausgelesen werden. Aufgrund der geringen Voraussetzungen arbeitet der dem Messmodell zugrunde liegende Algorithmus recht Ressourcen schonend. Es können in der Praxis mehrere tausende parallele Verbindungen mit einem Standard-PC beurteilt werden. Diese Meßmethode ist darüber hinaus in der Lage, aktuell ablaufende Gespräche zu beurteilen. Für die Messungen müssen keine Gespräche im Netzwerk simuliert werden. In einigen Netzwerken kann es jedoch passieren, dass die Simulation von zusätzlichen Gesprächen aufgrund einer zu geringer Bandbreite nicht möglich ist. In solchen Fällen ist man gezwungen die aktuell laufenden Gespräche zu analysieren.

Da sich das E-Modell auf die Paketparameter bezieht, können auch nur paketbezogene Fehler ausgewertet werden. Die Paketparameter repräsentieren jedoch nicht alle möglichen Fehlerquellen in der gesamten Ende-zu-Ende-Beziehung eines Telefonats. Die beurteilte Qualität verkörpert damit nur den QoS-Wert des lokalen Netzabschnitts.

Somit scheidet der vom E-Modell ermittelte MOS-Wert für die umfassende Qualitätsbeurteilung auf einer Ende-zu-Ende-Basis aus.

Objektive Bewertung der Sprachqualität

Ein objektives Messverfahren, das die zum subjektiven Qualitätseindruck führenden physiologischen und kognitiven Vorgänge simuliert, ersetzt heute vielfach die subjektiven Messverfahren. Die Qualität des zu bewertenden Testsignals wird anhand seiner Abweichungen von dem als Referenz dienenden Originalsignal bestimmt. Dazu werden beide Signale in eine gehörangepasste Darstellung umgeformt. Anschließend werden durch einen Vergleich im Zeit- und Frequenzbereich verschiedene Abstandsmaße bestimmt. Diese Ausgangswerte werden zu einer einzelnen Kenngröße zusammengefasst, der sich auf einen Schätzwert für die subjektiv empfundene Audioqualität abbilden lässt. Entsprechend der Bezeichnung der aus Hörtests gewonnenen Qualitätsbewertung als „Subjective Difference Grade“ (SDG) wird dieser Ausgangswert als „Objective Difference Grade“ (ODG) bezeichnet.

Es existieren zwei unterschiedliche Grundprinzipien, nach denen gehörangepasste Messverfahren arbeiten können:

- der Vergleich der Störung mit einer aus dem Originalsignal berechneten Maskierungsschwelle (Masked Threshold Concept) und
- der Vergleich zwischen gehörangepassten Signaldarstellungen von Originalsignal und zu bewertendem Signal (Comparison of Internal Representations).
- Als dritter Ansatz kann der direkte Vergleich der Spektraldarstellungen beider Signale (ohne Verwendung eines Gehörmodells) betrachtet werden.

Vergleich zwischen Maskierungsschwelle und Störung

Das Prinzip des Vergleichs der Störung mit einer Maskierungsschwelle (Masked Threshold Concept bzw. Noise Signal Evaluation) wurde bei den ersten bekannten gehörangepassten Messverfahren genutzt. Dabei wird durch Subtraktion des Originalsignals vom zu bewertenden Signal das Fehlersignal berechnet und mit einer aus dem Originalsignal bestimmten Maskierungsschwelle verglichen.

Die Vorteile dieses Konzeptes sind der relativ einfache Abgleich mit Hilfe der aus psychoakustischen Experimenten gefundener Daten und die Verwendbarkeit des zugehörigen psychoakustischen Modells für Audiocodierv Verfahren.

Vergleich zwischen gehörangepassten Signaldarstellungen

Das Konzept des Vergleichs zwischen gehörangepassten Signaldarstellungen (Comparison of Internal Representations) bildet die Grundlage für die meisten neueren gehörangepassten Messverfahren. Dabei werden sowohl aus dem Originalsignal als auch aus dem zu bewertenden Signal gehörangepasste Darstellungen (sogenannte Erregungsmuster) bestimmt. Die Bewertung der Qualität erfolgt aus dem Vergleich dieser Erregungsmuster.

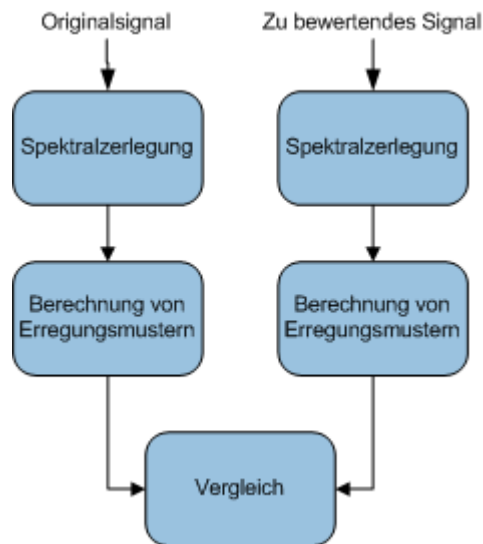


Abbildung 1: Vergleich zwischen gehörangepassten Signaldarstellungen

Einige Effekte, wie beispielsweise die Wahrnehmung einer Grundfrequenz in Tonkomplexen, lassen sich anhand linearer Spektraldarstellungen einfacher modellieren als anhand einer gehörrichtigen Darstellung. Ein solcher Ansatz kann zwar wegen des fehlenden Gehörmodells nicht die alleinige Grundlage eines Messverfahrens sein, er kann ein solches Modell aber ergänzen, da er zusätzliche Informationen über das Testsignal liefert, die aus einem Gehörmodell nur schwer gewonnen werden können.

Schallsignale werden durch Außen- und Mittelohr in ihrer Bandbreite beschränkt. In der Gehörschnecke erfolgt die Analyse der Schallsignale und anschließend die Umsetzung in Nervensignale. Ein Grundrauschen, welches durch die Strömung des Blutes im Kopf erzeugt wird, verdeckt sehr leise Schallsignale. Abhängig von der Frequenz wird die in der Gehörschnecke befindliche Basilarmembran an unterschiedlichen Orten maximal ausgelenkt. Die Haarzellen, welche sich auf der Basilarmembran befinden, werden durch diese Auslenkung erregt. Jede Haarzelle reagiert auf einen Bereich benachbarter Frequenzen. Ein konstanter Abstand auf der Basilarmembran ist eng verbunden mit der Wahrnehmung der Töne. Die Aufteilung des gesamten Hörbereichs erfolgt dabei nicht linear. Je nach psychoakustischem Experiment findet man unterschiedliche Übertragungsfunktionen von der Frequenzskala auf eine Tonskala. Die am häufigsten genutzte Skala ist die Bark-Skala. Diese umfasst einen Bereich von 0 Hz bis 15000 Hz und wird durch in 24 einander nicht überlappende Bereiche aufgeteilt.

Jede einzelne Haarzelle reagiert auf einen Bereich benachbarter Frequenzen. Dies kann als eine Bandpassfilterung betrachtet werden. Die Flankensteilheit der resultierenden Filter ist konstant, wenn man die Filterkurven in Pegelschreibweise über der Tonskala aufträgt. Die Steilheit der unteren Flanke der Filter ist unabhängig vom Pegel des erregenden Signals (ca. 27 dB/Bark). Die Steilheit der oberen Flanke ist für leise Signale betragsmäßig größer als für laute Signale (-30 dB/Bark bis -5 dB/Bark). Diese Pegelabhängigkeit der Steilheit wird durch neuronale Rückkoppelungsmechanismen beeinflusst. Diese Regelvorgänge benötigen eine Einschwingzeit. Die beste Frequenzselektivität wird deshalb erst einige Millisekunden nach Beginn eines Schallreizes erreicht.

Das Erregungsmuster eines Schallreizes, der aus mehreren Komponenten besteht, ergibt sich durch nichtlineare Addition der Erregungsmuster der Einzelkomponenten. Nach dem Ende eines Schallreizes benötigen die Haarzellen eine Erholungszeit bis sie ihre volle Empfindlichkeit wieder erreichen. Die Dauer dieser Erholungszeit ist abhängig von Pegel und Dauer des Schallreizes und kann bis zu mehreren hundert Millisekunden betragen. Signale mit hohem Pegel werden auf dem Weg zwischen Innenohr und Gehirn schneller verarbeitet als leise Signale. Ein lautes Signal kann dadurch ein vorhergehendes leises Signal unhörbar machen.

Die Erregungsmuster von Schallsignalen werden im menschlichen Gehirn weiterverarbeitet und gespeichert. Es werden dabei drei verschiedene Arten der Speicherung unterschieden: Langzeit-, Kurzzeit- und Ultra-Kurzzeitgedächtnis. Hörtests zur Bewertung von Audiosignalen basieren überwiegend auf dem Ultra-Kurzzeitgedächtnis, welches die meisten Feinheiten eines Signals speichern kann. Seine Dauer ist allerdings auf Signalabschnitte mit einer maximalen Dauer von fünf bis acht Sekunden beschränkt. Extrem kleine Unterschiede zwischen Signalen werden nicht wahrgenommen. Die Wahrnehmungsschwelle ist dadurch definiert, dass die Wahrscheinlichkeit der Wahrnehmung an dieser Stelle 50 Prozent beträgt.

Ein in vollständiger Ruhe deutlich wahrnehmbares Signal kann in Anwesenheit eines zweiten Signals unhörbar, d. h. verdeckt werden. Dieses zweite Signal wird dann Maskierer genannt. Bei der Simultanverdeckung sind beide Signale gleichzeitig vorhanden. Die Stärke der Maskierung hängt hauptsächlich von Struktur, Pegel und relativer Frequenzlage, teilweise aber auch von der absoluten Frequenz der Signale ab. Die maximale Verdeckung wird jeweils erzielt wenn Maskierer und verdecktes Signal die gleiche Mittenfrequenz haben. Bei abweichender Mittenfrequenz ergibt sich ein der Erregung entsprechender Verlauf. Die maximale Verdeckung liegt bei der Verdeckung von Tönen durch Rauschen unabhängig von der Mittenfrequenz des Rauschens ca. 5 dB unter dem Pegel des Maskierers. Eine zeitliche Verdeckung tritt auf, wenn Maskierer und verdecktes Signal nicht gleichzeitig vorhanden sind. Entsprechend der Erholungszeit bei den Erregungsmustern werden leise Signale kurz nach lauten Signalen durch diese unhörbar. Laute Signale nach leisen Signalen überholen diese bei der Weiterleitung vom Ohr zum Gehirn und verdecken diese.

Die wahrgenommen Lautstärke eines Audiosignals hängt außer vom Schalldruckpegel auch von seiner Frequenz und Dauer ab. Anteile eines komplexen Schallereignisses können sich gegenseitig verdecken, so dass die resultierende Lautstärke geringer sein kann als die Summe der Lautstärke der Einzelschalle. Für die Beurteilung von Audiosignalen im Vergleich zu einem Referenzsignal ist die Lautstärke der Störkomponenten wichtig. Die Lautstärke dieser Störkomponenten ist auf Grund der Drosselung durch das Nutzsignal reduziert. Der Parameter „Schärfe“ beurteilt die Klangfarbe eines Audiosignals. Ein Klang wird als „scharf“ wahrgenommen, wenn er viele hochfrequente Komponenten enthält.

Objektive Methoden zur Messung der Einwege-Sprachqualität

Die momentan verfügbaren und international anerkannten objektiven Messszenarien zur Qualitätsbeurteilung der Sprache basieren auf den Einwege-Verfahren. Im Laufe der Jahre wurden von diversen Herstellern und Telekommunikationsorganisationen eine Reihe unterschiedlicher proprietärer Algorithmen zur adäquaten Qualitätsbeurteilung der Einwege-Übertragungsqualität entwickelt. Zu diesen Algorithmen gehören: TOSQA (Telecommunications Objective Speech Quality Assessment), PAMS (Perceptual Analysis/Masurement System), PSQM (Perceptual Speech Quality Measurement), PSQM+. Diese objektiven Messmechanismen vergleichen einen übermit-

telten Sprachstrom auf die während des Transports aufgetretenen Qualitätsveränderungen. In den vergangenen zehn Jahren haben die Telekommunikationsgremien intensiv an der Validierung der jeweiligen Messverfahren und -Algorithmen gearbeitet und abschließend in Form der ITU-T Empfehlung P.862 eine objektive Methode zur Ende-zu-Ende-Sprachqualitätsbeurteilung für schmalbandige Telefonnetze und Sprach-Codecs veröffentlicht. Die ITU-T Empfehlung P.862 basiert auf der Harmonisierung der zwei konkurrierenden Algorithmen PAMS99 und PSQM+. Dieser neue Algorithmus wurde Perceptual Evaluation of Speech Quality (PESQ) genannt.

Objektive Bewertung der Sprachqualität nach PESQ

PESQ (Perceptual evaluation of speech quality) ist in der ITU-Spezifikation P.862 veröffentlicht. Dieser Standard dient der Überprüfung der Sprachqualität von Telefongesprächen bis 3.2 KHz. Außerdem gewährleistet der Standard eine echte Ende-zu-Ende-Messung. Der PESQ-Algorithmus erkennt darüber hinaus variable Jitter-Werte, Paketverluste, Verzerrungen und Übertragungsfehler.

Als Ausgangspunkt einer PESQ-Messung wird ein Referenzsignal genutzt. Dieses wird vom Messgerät durch den "Prüfling" (beispielsweise das VoIP-Netzwerk) geschickt. Beim Empfänger wird das durch die Übertragung verminderte Signal aufgezeichnet und durch den PESQ-Algorithmus mit dem ursprünglichen Referenzsignal verglichen.

Bei dem Referenzsignal handelt es sich um eine von der ITU vorgegebene Wave-Datei mit folgenden Eigenschaften: 16 Bit, Mono, PCM, 8 KHz. Diese Wave-Datei enthält einen Monolog, welcher die vorgegebenen Voraussetzungen erfüllt. Für eine Messung bei VoIP wird dieses Signal durch einen entsprechenden Sprach-Codec kodiert. Nach der Kodierung wird das kodierte Referenzfile in RTP-Pakete verpackt und durch das Netzwerk zum anderen Ende der Telefonverbindung übermittelt. Bei dem Durchlaufen des Netzwerkes können mehrere Faktoren auf das Referenzfile einwirken und dieses verändern. Auf der Empfängerseite werden die Pakete anschließend gespiegelt und an den ursprünglichen Sender wieder zurückgeschickt. Der ursprüngliche Sender (und somit der Empfänger des gespiegelten Signalstroms) empfängt die Daten und dekodiert diese. Die dekodierte Wave-Datei wird anschließend mit der Original-Wave-Datei dem PESQ-Algorithmus zur Verfügung gestellt und aus dem Vergleich der notwendige PESQ-Wert errechnet. PESQ hat einen Wertebereich zwischen -0,5 und 4, 5.

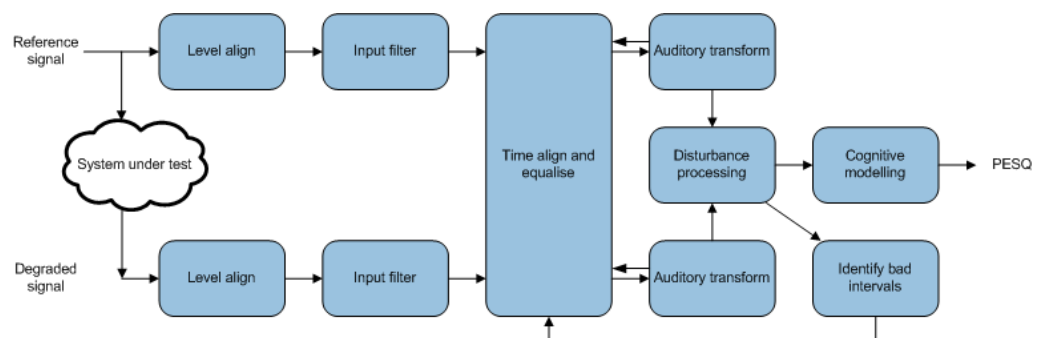


Abbildung 2: Blockschaltbild des PESQ-Algorithmus

Wie im Blockschaltbild dargestellt, beginnt der Messprozess mit dem Referenzsignal. Dieses wird durch den "Prüfling" (Wolke) zum anderen Ende der Teststrecke übermittelt. Das Referenzsignal und das verminderte Signal, welches durch den Prüfling geschickt wurde, durchlaufen während der Messung verschiedene Filter. Hierbei werden Zeitabgleiche und Korrelationen durchgeführt, um die Veränderungen des verminderten Signals zu berechnen. Diese Verfahren ermöglichen es, eine Verzögerung in einem Signal zu lokalisieren. Die Signale werden beim Empfänger in ein internes Format gewandelt und miteinander verglichen. Ein kognitives Modell berechnet anschließend den PESQ-Wert. Ein Psycho-akustisches-Gehörmodell sorgt bei diesen Mess- und Vergleichswerten für eine Nachbildung des menschlichen Hörempfindens und berücksichtigt somit rechnerisch die subjektiven Beurteilungskriterien von Menschen.

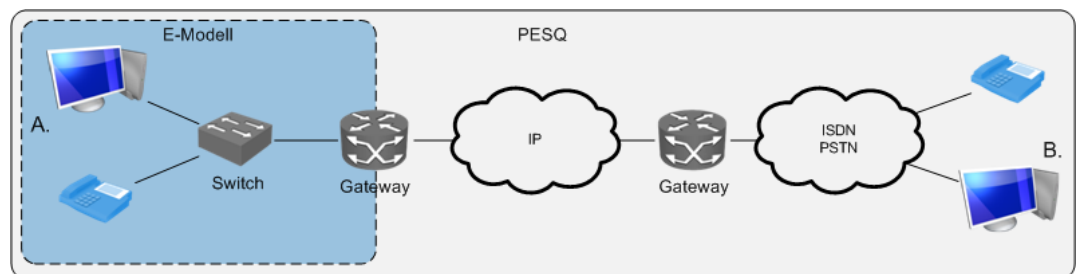


Abbildung 1: Beispiel einer PESQ-Messung im Netzwerk

VoIP Gateways

Mit VoIP Gateways werden zwei unterschiedliche Telefontechnologien aneinander angepasst. So wird beispielsweise zwischen der klassischen Telefon und der VoIP-Welt eine VoIP/TDM-Umwandlung vorgenommen. Gateways übersetzen auch zwischen zwei unterschiedlichen Sprachcodierungen wie beispielsweise G.711/G.729. Bei den Umsetzungen in das jeweilige andere Format werden aus den VoIP-Paketen die RTP-Streams mit den zugehörigen Sprachinformationen entpackt und für die andere Technologie neu zusammengestellt und in das Netz übertragen.

VoIP Gateways finden ebenfalls Ihren Einsatz, wenn Gespräche über das WAN geführt werden sollen und hierfür eine geringe Bandbreite zur Verfügung steht. Innerhalb des Unternehmens werden die Gespräche auf Basis des G.711 Codecs geführt. Für den Transport über die WAN-Strecke werden die Sprachströme in ein G.729 Format gewandelt. Durch die Umwandlung auf den G.729 Codec wird auf der WAN-Strecke erheblich Bandbreite eingespart. Nachteilig wirkt sich jedoch die Verringerung der Gesprächsqualität aus.

Reduktion der Sprachqualität bei der Übermittlung über VoIP Gateways

Im Grunde genommen kann die Sprachqualität eines Gateways nur mit Hilfe einer Gruppe menschlicher Zuhörer bewertet werden. Diese Testprozedur ist jedoch teuer und zeitaufwändig. Daher entwickelt die Industrie die hierfür notwendigen objektiven Messwerkzeuge. Die Messwerkzeuge der Nextragen GmbH basieren auf einer vernünftigen Kombination von subjektiven und objektiven Messmethoden und sind dadurch in der Lage, die Sprachqualität auf einer Ende-zu-Ende-Basis zu beurteilen. Dadurch fließt auch in die Messergebnisse die Güte der im Datenpfad vorhandenen Gateways ein.

Die Nextragen Lösung hat den am besten geeigneten Testalgorithmus (PESQ) in Ihren Produkten implementiert. Die Kombination aus objektiven und subjektiven Tests sorgt für eine sehr präzise Beurteilung der Sprachqualität auf einer Ende-zu-Ende-Basis und umfasst alle Koppelkomponenten und Gateways im Übermittlungsnetz.

Warum sind Sprachqualitätsmessungen bei VoIP Gateways für den Netzbetrieb wichtig?

Ein VoIP-Netzwerk eines Unternehmens ist über ein VoIP-Gateway mit dem öffentlichen ISDN/PSTN-Netz verbunden. Alle Gespräche zwischen den Telefonnutzern am Unternehmensnetzwerk und Teilnehmern des öffentlichen Telefonnetzes (und umgekehrt) werden über das VoIP-Gateway geführt. Überprüft man die Sprachqualität einer solchen Netzverminderung mit Hilfe des E-Modells, dann kontrolliert man nur die IP-Netzstrecke zum Gateway und nicht die Eigenschaften des eigentlichen Gateways. Bei einer Überprüfung auf Basis des PESQ-Algorithmuses wird ein Signal durch das Gateway geschleust und die Sprachqualität, quasi als Charakteristik der Gateway-Performance, ermittelt.

Untersucht der Messmechanismus nur die Sprachqualität für den Netzabschnitt bis zum ersten Gateway (Umsetzer zwischen dem Unternehmensnetz und den öffentlichen Fernsprechressourcen) stehen bei der Nutzung des Messverfahrens nach dem E-Modell die für die Messung notwendigen RTP-Informationen für die Berechnung der Sprachqualität nach dem Gateway nicht mehr zur Verfügung. Folglich kann mit dem E-Modell nur eine qualifizierte Aussage des QoS bis zum Gateway getroffen werden. Fehler hinter dem Gateway bleiben für das E-Modell verborgen. Da bei einer Messung über mehrere Netzelemente nicht sichergestellt werden kann, dass alle notwendigen Paketparameter durchgängig zur Verfügung stehen, eignet sich das E-Modell nicht für eine Ende-zu-Ende Bewertung von Sprachqualitäten bei VoIP.

Nur der PESQ-Algorithmus überwindet diese Hürde, und kann die gesamte Strecke (Ende-zu-Ende) über Gateways hinweg beurteilen. Dadurch werden bei der PESQ-Methode sämtliche auftretenden Fehler auf der gesamten Übertragungsstrecke sichtbar gemacht. Hierzu gehören beispielsweise:

- Die in IP-Netzen üblichen Fehler wie beispielsweise Paketverlusten und Jitter
- Die durch die Codec-Konvertierung verursachten Leistungsminderungen (z.B. Veränderungen des Frequenzspektrums)
- Formatumsetzungsfehler bei der Übersetzung zwischen dem IP- und dem öffentlichen ISDN/PSTN Netz
- Netzspezifische Fehler bzw. Übertragungsfehler im ISDN/PSTN, die sich durch eine erhöhte Bit Error Rate (BER) äußern
- Fehler bei der Umsetzung von Codecs
- Fehler die durch eine Überlastung vom Gateway entstehen können

Fazit

Sprachqualitätsmessungen in heutigen IT-Infrastrukturen können durch die Komplexität der hybriden Telekommunikationslösungen zu einem schwierigen Unterfangen werden. Die richtige Vorbereitung und das notwendige Know-how legen die Basis für die gewünschten Messerfolge. Die Messtechnik der Nextragen GmbH liefert die Werkzeuge, um die Probleme in heutigen Kommunikationslandschaften zu ermitteln.

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.



Nextragen GmbH
Lise-Meitner-Str.2
24941 Flensburg

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

www.nextragen.de
info@nextragen.de

Änderungen und Irrtümer vorbehalten