Audioqualität in der VoIP-Telefonie

Mit dem Verstehen und Verstanden werden beginnt eine gute Kommunikation. Dies gilt natürlich in besonderen Maße für jede Art des verbalen Austausches – denn im Gegensatz zum geschriebenen Gedankenaustausch kann man hier nicht noch einmal nachlesen, sondern muss im schlimmsten Fall nachfragen.

Ist der Anruf abgehackt oder verzerrt ist dieser Informationsfluss gestört und sorgt schnell für Unmut. Schon seit dem Anbeginn der Telekommunikation war es daher wichtig, eine möglichst unterbrechungsfreie, Störungsunabhängige Echtzeit-Verbindung zu schaffen – damals noch per Kupferdraht, heute über das Internet.

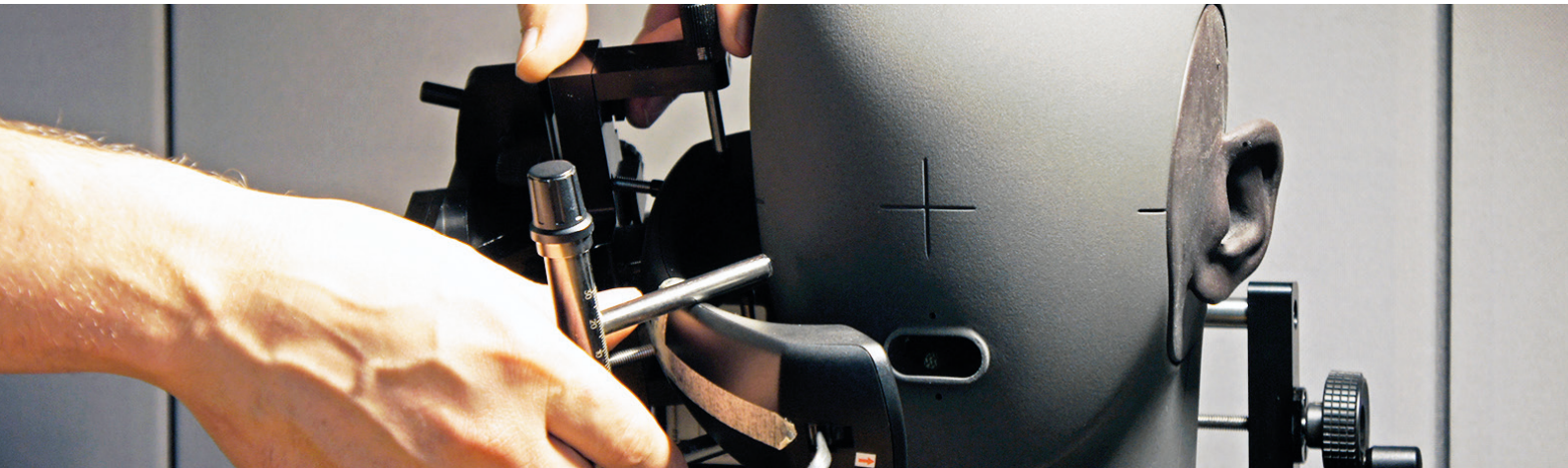
Und das ist gerade bei der Telekommunikation nicht so einfach, denn es muss sichergestellt werden, dass die Datenpakete in der richtigen Reihenfolge und zur richtigen Zeit beim Empfänger ankommen. Je nach Internetverbindung, Verbindungsqualität, Netzwerkrouting und Land des Senders oder Empfängers kann es schnell zu Problemen kommen, welche die Tonqualität beeinflussen. Aber nicht nur die Technik.

Soweit so gut – dank der heutigen Technik sowie fortschreitenden Digitalisierung sollte es doch kein Problem mehr sein, glasklare Audio- und Videokommunikation über das Internet anzubieten. Schließlich macht die Technik jedes Jahr große Fortschritte und auch schnelle Internetanbindungen stehen für geschäftliche oder private Nutzer flächendeckend zur Verfügung. Weit gefehlt – hinter einer glasklaren, unterbrechungsfreien Kommunikation steht sowohl komplizierte Technik, als auch viel Knowhow.

Snom hat sich in den über 20 Jahren seines Bestehens immer sehr große Gedanken im Bereich Qualität und Qualitätsoptimierung gemacht. Schon früh wurde hier in das Wissen um gute Audioqualität investiert – vom eigenen Audiolabor über extrem aufwendige Tests. Im Folgenden möchten wir Ihnen ein wenig Einblick in diese Welt vermitteln und Ihnen erklären, warum bei Snom das Audio immer ein wenig bis extrem viel besser ist als bei den meisten.

Immer im Zentrum – der Codec:

Wenn man über die Audioqualität im Bereich VoIP redet, dann muss man über Codecs sprechen. Mit Hilfe eines Audio-Codecs werden die Signale eines Anrufes digitalisiert, kodiert, an der Nebenstelle entsprechend dekodiert und wieder in ein analoges Signal umgewandelt. Je nachdem welche lokalen Gegebenheiten vorherrschen, muss der Codec dabei dynamisch und möglichst schnell das Audiosignal bearbeiten und weiterversenden. Das wäre an sich kein Problem, hätte man alle Zeit der Welt, aber gerade in der Telefonie muss die Kommunikation möglichst in Echtzeit stattfinden. Ansonsten kommt es zu den Effekten, die der Leser noch aus



den analogen Telefonaten in weit entfernte Länder kennt: man hört die Gegenseite immer mit einer deutlichen Zeitverzögerung.

Die Aufgabe des Codes ist es also auch, die Daten während der Digitalisierung möglichst verlustfrei zu optimieren, damit die Datenpakete möglichst effizient und ohne hohe Bandbreitennutzung schnell beim Empfänger landen und dort auch schnell wieder dekodiert werden können.

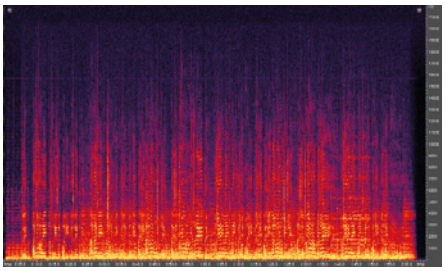


Abbildung 1.
Unkomprimiertes Audiospektrum

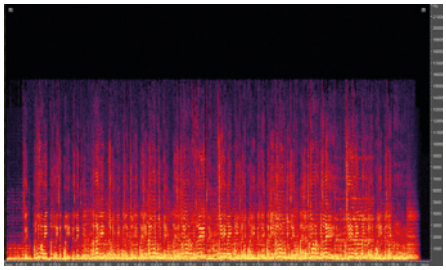


Abbildung 2. Mit MP3 komprimiertes Audiospektrum mit begrenzte Bandbreite

Verdeutlichen kann man diesen Vorgang mit der Einführung des MP3 Formats in den 90er Jahren. Das damals neuartige Verfahren machte sich eine Eigenschaft des menschlichen Gehöres zu Nutzen. Ein Mensch kann zwei unterschiedliche Töne erst dann auseinanderhalten, wenn ein gewisser Mindestunterschied der Tonhöhen vorliegt. Gerade vor oder nach lauten Tönen kann das menschliche Gehör leisere Töne schlechter und gar nicht mehr wahrnehmen. Das Prinzip das unhörbare Teile eines Musikstückes nicht abgespeichert werden müssen machte sich das MP3-Format zu nutzen und schaffte es so, sehr große Audiodateien deutlich effizienter abzuspeichern. Und bei der Telefonie? An sich arbeiten Codecs in der VoIP-Telefonie ähnlich. Innerhalb des Kodiervorganges wird das analoge Signal digitalisiert und komprimiert um es für den unterbrechungsfreien Versand vorzubereiten. Diese Kompression findet nicht verlustfrei statt, kann aber auf Bereiche reduziert werden, die für das menschliche Ohr kaum wahrnehmbar sind. Gleichzeitig sollen die einzelnen Datenpakete möglichst klein gehalten werden um möglichst unterbrechungsfrei und schnell verschickt werden zu können – denn die Bandbreite oder Qualität der Internetanbindung oder des lokalen Netzwerkes kann variieren.

Heutzutage werden in der internationalen Festnetz- und Mobiltelefonie hauptsächlich von der Internationalen Fernmeldeunion standardisierte Codecs wie der G.711, G.722, AMR, iLBC und Opus verwendet.
G.711 Codec – Meist verbreitet, aber Datenhungrig
Der G.711 Standard wird heutzutage in dem meisten Festnetz und IP-Telefonen genutzt, da er eine gute Sprachqualität bei gleichzeitiger Kompression ermöglicht.

Der Codec arbeitet dabei wie folgt:
Alle 125 Mikrosekunden erstellt G.711 bei einer Abtastrate von 8.000 Herz ein Sample des aufgenommenen Audiosignals. Dieses Sample wird dann durch den Codec auf 8bit komprimiert. Um das Signal gleichzeitig zu optimieren und Bandbreite zu sparen begrenzt G.711 nun den Frequenzbereich von ursprünglich ca. 15.000 Herz auf 300 bis 3400 Hz. Das spart zum einen Speicherplatz bzw. Bandbreite und optimiert so das Signal schränkt aber auch die Dynamik des Signals ein.

Trotz dieses eingeschränkten Frequenzbereichs von ursprünglich 8000Herz erreicht der G.711 Codec bei so genannten Mean Opinion Score (MOS) Messung einen Wert von 4,4 von 5 Punkten (wenn wir nur mit anderen schmalband Codecs vergleichen). Der MOS ermittelt dabei das subjektive Empfinden der Sprachqualität eines Nutzers. Der G.711 Codec nutzt in Europa weit verbreitete A-law Verfahren um die bei der Komprimierung verlorene Dynamik wiederherzustellen. Dabei wird das Signal bei der Digitalisierung mit einer nichtlinearen Kennlinie (A-Kennlinie) quantisiert. Bei der Quantisierung wird ein analoges Audiosignal digitalisiert – also umgewandelt.

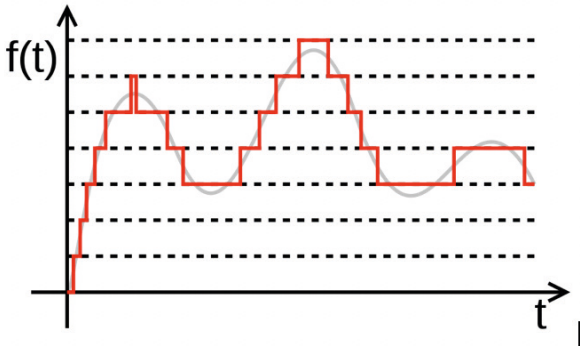


Abbildung 3. Ein Beispiel für die Quantisierung eines Audiosignals. Rot gekennzeichnet ist die spätere digitale Form (hier ohne Anwendung von der A-Kennlinie) (Quelle: Wikipedia)

Das Ziel ist es, den Dynamikumfang zu erhöhen und gleichzeitig das Signal-Rausch-Verhältnis zu vergrößern. Um das zu erreichen werden bei dieser logarithmischen Quantisierungskennlinie große Signalauslenkungen feiner und kleine Signalauslenkungen gröber aufgelöst.

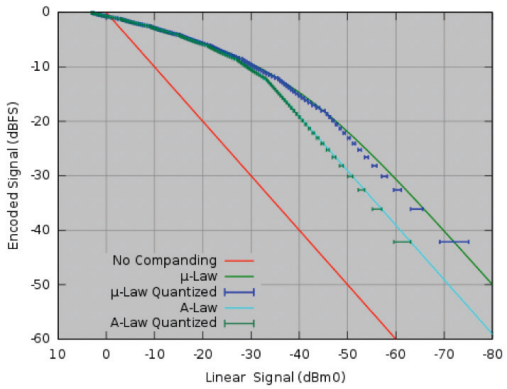


Abbildung 4. Die A- und µ-Kennlinie

Diese hohe Sprachqualität hat allerdings auch ihren Preis. Inklusive des Overheads, also dem Versenden von nicht Nutzerrelevanten Daten benötigt der G.711 Codecs Datenübertragungsraten von 80kbit/s bis 128 kbit/s. Andere Codecs sind hier deutlich sparsamer wie beispielsweise der Opus Codec, der mit 20-48 kbit/s auskommt.

Es bedarf aber mehr um eine saubere und vor allem klare IP-Kommunikation gewährleisten zu können. Ist das ehemalige analoge Signal einmal digitalisiert muss es vom Sender zum Empfänger transportiert werden. Dafür werden verschiedene Protokolle genutzt die besonderen Eigenschaften vorweisen müssen um Echtzeitkommunikation zu ermöglichen.

Protokolle, vor allem Internet Protokolle stellen die Grundlage des Internets dar. Die Protokolle verbinden dabei die verschiedenen Vermittlungsschichten innerhalb von Netzwerken und dem Internet nach dem OSI-Modell. Wichtig ist, dass je nach Netzwerk bzw. Schicht Sender und Empfänger nach festgelegten Regeln arbeiten müssen. Diese Regeln werden innerhalb der Protokolle festgelegt

Das SIP-Protokoll:

Ganz am Anfang steht das SIP, das Session Initiation Protocol. SIP ist ein Netzprotokoll das den Aufbau, die Steuerung und die Kommunikationssitzung zwischen zwei Teilnehmern steuert und verwaltet. Das SIP-Protokoll handelt also die Kommunikationsmodalitäten zwischen den Teilnehmern aus. Der eigentliche Datenverkehr wird über andere Protokolle durchgeführt, die teilweise in SIP eingebettet werden wie beispielsweise das RTP- oder SDP-Protokoll.

Teilnehmern wird bei SIP eine eindeutige Adresse zugeordnet, ähnlich einer E-Mail-Adresse. Diese Adresse enthält einen eindeutigen Benutzernamen sowie eine Domain. Wird eine Session per SIP zwischen zwei Endgeräten aufgebaut tauschen die Geräte erst einmal alle relevanten Informationen über Eigenschaften und Fähigkeiten aus. Wird ein Telefongespräch aufgebaut werden Daten in Echtzeit und direkt zwischen den Geräten ausgetauscht. Hierzu wird das Real-Time Transport Protocol (RTP) eingesetzt.

Das RTP-Protokoll:

Mit dem RTP Protokoll wird für eine kontinuierliche Übertragung von audiovisuellen Daten sichergestellt. Dabei werden Daten entsprechend in Echtzeit effizient kodiert und paketiert – ähnlich wie bei den vorher besprochenen Audiocodecs. Eine weitere Form des RTP-Protokolls ist das SRTP (Secure Real-Time Transport Protokoll) eine verschlüsselte Variante von RTP.

Wichtig ist, dass weder RTP noch SRTP für die Übertragen der Daten verantwortlich ist – das übernimmt das in das RTP integrierte UDP-Protokoll.

UDP Protokoll – Schnell aber unzuverlässig

Kommunikation findet immer in Echtzeit statt, daher muss die gesamte Datenkommunikation möglichst schnell und effektiv stattfinden. Das UDP Protokoll ist dabei sehr „minimalistisch" ausgelegt und dadurch nicht auf Zuverlässigkeit ausgelegt. Das bedeutet, dass beispielsweise keine Empfangsbestätigungen ausgetauscht werden, wenn ein Paket übertragen wurde. Dadurch besteht keine Übertragungsgarantie und somit können fehlerhaft Übertragungen nicht ausgeschlossen werden. Bei UDP werden Daten also „just in time" ausgeliefert – ob sie erfolgreich ankommen oder nicht. Wichtig ist aber nicht die zu 100% erfolgreiche Übertragung der Datenpakete sondern eine möglichst geringe Laufzeit der Pakete.

Um Voice over IP Telefonie zu realisieren, werden zwei grundlegende Protokolle angewendet. Das SIP-Protokoll (Session Initiation Protocol) baut dabei die Verbindung von einem Teilnehmer zum anderen auf. Es stellt, sozusagen, die Weichen für die Daten. Ist die Verbindung aufgebaut, wird RTP (Real time Transport Protocol) genutzt um die Audio- und Videoströme zu übertragen. Um sichere Gespräche führen zu können wurde eine verschlüsselte Variante dieser beiden Protokolle entwickelt. So entstanden SIP und SRTP. Über diese Protokolle wird der Verbindungsaufbau zwischen IP-Telefonanlage und VoIP-Telefon, mit Hilfe des sog „Handshake-Verfahrens", verschlüsselt durchgeführt und kann so zwar noch mitgeschnitten, aber nicht mehr ausgelesen werden.

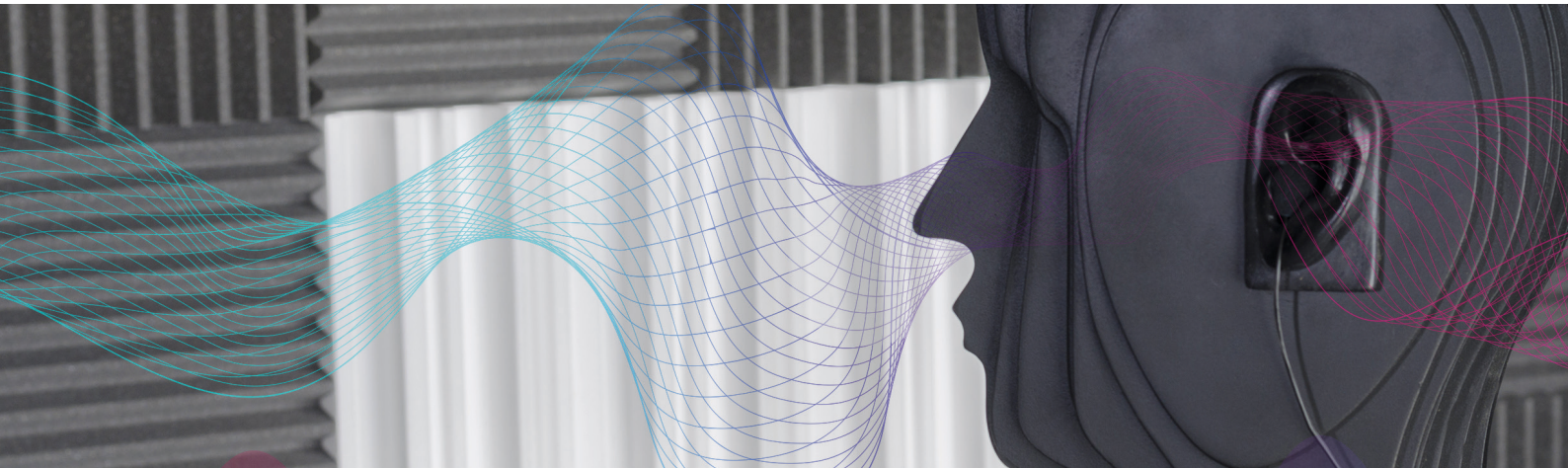
Ist die Verbindung aufgebaut, kommt das SRTP zum Tragen, welches die Sprache kodiert, in verschlüsselte Datenpakete verpackt und vom Sender (IP-Telefonanlage), über das Internet, zum Empfänger versendet. Damit dieser Empfänger die Daten auch wieder entschlüsseln kann, wird bei der Initialisierung der Verbindung ein Master-Key, über SIPS mitgesendet, der dem Empfänger ermöglicht, die verschlüsselten Daten wieder "hörbar" zu machen.

Mögliche Störfaktoren:

Routing und VLAN: Wir schon erwähnt wird in der digitalen Kommunikation das gesprochene Wort durch Codecs digitalisiert und in Datenpakete komprimiert. Damit ist es aber noch nicht getan – denn der Weg den das Datenpaket nun durch das Netzwerk oder das Internet nimmt, ist entscheidend. Im richtigen Moment müssen alle wichtigen Fragmente des Datenpaketes am Zielort eintreffen um wieder zusammengesetzt werden. Ansonsten kommt es zum so genannten „Jitter", einer Varianz der Laufzeit von Datenpaketen, hörbar als „zerhackte" oder „metallische" Stimme.

Daten trennen:

Ein weiterer wichtiger Punkt ist die Trennung von Audio- Datenpaketen und dem Datenfluss beim „normalen" Surfen im Internet oder der Übertragung einer E-Mail. Werden beide Paketarten gleichzeitig in einem Netzwerk übertragen, können sich diese Pakete gegenseitig behindern bzw. Pakete zu früh, zu spät oder gar nicht beim Empfänger ankommen. Eine erhöhte Latenz oder weiterer „Jitter" wäre das Resultat. Der Grund dafür ist die Art der Protokolle der IP-Telefonie. Zum Einsatz kommen das so genannte RTP (Real-Time Transport Protocol), SRTP (Real-Time-Control-Protocol) und das



User Control Protocol (UDP). Gerade letzteres Protokoll wird aufgrund seiner niedrigen Latenz benötigt, da es keine Rückmeldung über den erfolgreichen Empfang eines Datenpaketes gibt. Es ist also im Vergleich zum bekannten TCP Protokoll nicht auf Zuverlässigkeit, sondern Geschwindigkeit ausgelegt. Dafür kann das UDP Protokoll durch eine sehr geringe Latenzzeit punkten. Die Aufgabe des RTP bzw. SRTP Protokolls ist es wiederum, Datenströme die in Echtzeit benötigt werden zu übertragen, zu steuern und zu kontrollieren.

Quality of Service:
Aber auch das Optimieren der QoS (Quality of Service) Einstellungen kann schon Abhilfe schaffen. Diese Einstellungen ermöglichen es dem Router den Netzwerkverkehr entsprechend der Nutzung zu priorisieren, also Sprachpaketen den Vorrang vor Datenpaketen einzuräumen.

Konnektivität:
Und zu guter Letzt liegt einer der wichtigsten Störfaktoren natürlich in der eigentlichen Internetanbindung. Je nach Größe der Installation bzw. Anzahl der angeschlossenen IP-Telefone wird eine entsprechend leistungsfähige Internetanbindung benötigt. Denn die fertigen Datenpakete müssen natürlich entsprechend hoch oder heruntergeladen werden. Je nach verwendetem Codec werden pro Anschluss zwischen 3Kbit/s (GSM) – bis zu 128kbit/s (G711 – G722) benötigt. Bei 2-5 Telefonen ist das erst mal kein großes Problem aber bei Unternehmen mit 50, 100 oder mehr Telefonen gerät eine einfache Internetanbindung schnell an ihre Grenzen.

Der kleine Unterschied
In der Echtzeitkommunikation sind viele Dinge zu beachten will man eine möglichst unterbrechungsfreie und hochwertige Audioverbindung herstellen. Neben der richtigen Hardware muss auch die genutzte Software großes leisten. Wie kann ein Hersteller von VoIP-Telefonen gewährleisten in jedem Moment und an jedem Ort immer die bestmöglichste Telefonverbindung herstellen zu können?

Das macht Snom:
VoIP-Telefonie benötigt ein breites Spektrum an technischen Voraussetzungen um sicher und vor allem verzögerungsfrei funktionieren zu können. Damit diese Prozesse im Telefon überhaupt stattfinden können spielt die verbaute Hardware natürlich eine entscheidende Rolle. Hier trennt sich die Spreu vom Weizen beim Einsatz hochentwickelter und vor allem leistungsfähiger Komponenten:

- Hardware zur Tonwiedergabe bzw. Aufnahme:
Die beste Internetanbindung, das schnellste Netzwerk kann nicht die ursprüngliche Aufnahme bzw. Wiedergabe am Telefonhörer oder am Lautsprecher im Freisprechmodus ersetzen. Hier gibt es gleich mehrere Punkte auf die geachtet werden muss. Im Telefonhörer sind Mikrofon und Lautsprecher integriert was andere Anforderungen als beispielsweise beim Freisprechen an sie stellt. Denn der Hörer wird während des Telefonats direkt an das Ohr bzw. den Mund geführt. In diesem extremen Nahbereich muss also die Empfindlichkeit des Mikrofons entsprechend „feingetuned“ werden, genauso wie die Ausgabe über den integrierten Lautsprecher.

Völlig anders verhält es sich beim zweiten Lautsprecher der für die Freisprecheinrichtung genutzt wird. Der Lautsprecher für die Freisprechanlage muss entsprechend laut sein ohne aber blechern zu klingen, das zweite Mikrofon für die Freisprechanlage muss den Sprecher einfangen können ohne zu viel Hall mit aufzuzeichnen und vieles mehr. Eine weitere Herausforderung beim Freisprechen ist das Echo das auftreten kann, wenn der durch den Raum zurückgeworfene Ton gleich mehrfach bzw. verzögert vom Mikrofon aufgefangen wird. Das Mikrofon und der Lautsprecher zum Freisprechen können sich also beeinflussen. Das kann entweder dazu führen, dass die Teilnehmer sich nicht mehr verstehen oder der sprechende Teilnehmer verwirrt wird, da er seine eigene Stimme hört.
Hier arbeiten unsere Experten mit komplizierten softwareseitigen Echokompensatoren die dieses Problem vermindern oder kompensieren können.
Um sicher zu stellen, dass die Qualität der verbau-

ten Teile stimmt, das Design der Telefone nicht die Aufnahme oder Wiedergabe behindert und die Sprachaufnahmen gut abgemischt sind, geht Snom noch einen Schritt weiter: In unserem Hauseigenen Audiolabor arbeiten Experten daran, unter Laborbedingungen den besten Sound aus jedem Snom Telefon herauszuholen. Und das für alle Gelegenheiten und Umgebungen. Im Einzel- oder Großraumbüro, in der Lagerhalle oder im Krankenhausflur. Kommunikation muss immer funktionieren. Das wird möglich durch umfangreiche Tests und den Einsatz von ausgeklügelten so genannten Equalizern.
Die Aufgabe eines Equalizers ist es, einzelne Frequenzen zu trennen um die Aufnahme oder Wiedergabe zu optimieren und etwa Sprache deutlicher hervorzuheben. Alles das geschieht natürlich in Echtzeit.

- Software oder Firmware:
Als letztes ist natürlich auch die Firmware auf dem Telefon selbst entscheidend. Die Firmware agiert als übergreifendes Kontrollorgan des Telefons und steuert alle Hard- und Software Komponenten. Sie muss darauf reagieren ob ein USB-Gerät angeschlossen wird, die Kommunikation mit der VoIP-Telefonanlage übernehmen, das User-Interface generieren und natürlich auch den Telefonanruf in Echtzeit unterstützen. Denn sobald eine Verbindung zwischen zwei Telefonen hergestellt wurde, arbeiten die Telefone direkt miteinander und müssen die Kodierung und De-Kodierung der Datenpakete übernehmen. Die Firmware muss also vor allem eines – vielfältige Aufgaben in Echtzeit übernehmen. Gerade das Zusammenspiel zwischen Telefonanlage und Telefon muss möglichst einfach und effizient geschehen, genauso wie die Verarbeitung der Menüs die auf dem Display angezeigt werden. Denn es muss immer genug Rechenleistung übrig bleiben um die Echtzeitkommunikation betreiben zu können. Snom arbeitet bereits seit über 20 Jahren an einer komplett selbstentwickelten, hocheffizienten und vor allem sehr sicheren Firmware, die alle diese Aufgaben problemlos übernehmen kann.

Soft- und Hardware sind Komponenten auf die Snom als Hersteller direkten Einfluss hat, aber es gibt auch externe Einflüsse die zu berücksichtigen sind um eine möglichst gute Sprachqualität gewährleisten zu können:

Zusammenfassung:

Moderne Echtzeitkommunikation ist für uns alle zum Alltag geworden – allerdings vergisst man schnell welcher technischer Aufwand und wie viel Erfahrung benötigt wird um sie zu ermöglichen. Alles beginnt schon beim ersten Design der Telefone und der optimalen Positionierung der Lautsprecher und Mikrofone. Im nächsten Schritt muss die Auswahl der genutzten Hardware stimmen, damit alle Komponenten nicht nur miteinander zusammenarbeiten können, sondern auch entsprechend Leistung vorhanden ist um in Echtzeit Daten effizient kodieren und versenden zu können.

Eine moderne Firmware wiederum sorgt dafür das die Komponenten richtig zusammenarbeiten können. Sie beinhaltet auch die entsprechenden Codecs und Protokolle die die aufgenommene Sprache digitalisieren, optimieren und schließlich über das Internet versenden. Ein optimiertes Netzwerk vor Ort bzw. die richtige Konfiguration des Routers stellt sicher, dass die generierten Datenpakete in Echtzeit übermittelt werden können.

Diese Komponenten müssen perfekt zusammen arbeiten um Telefonie in bester Ton- und Sprachqualität zu ermöglichen. Und das überall auf der Welt und innerhalb von möglichst jeder Infrastruktur. Hier spielt vor allem eines eine Rolle – Erfahrung und der stete Hang zur Perfektion!