

Codecs und Audiotbearbeitung

Inhalt

Codecs und Audiotbearbeitung	1
Codecs	2
PCM	2
ADPCM.....	2
G.711	2
G.722	3
G.726	3
Sonstige Codecs.....	3
Latenz und MP3.....	3
MP3 und ähnliche.....	4
Codecvergleich	5
Aushandlung.....	5
Speicherung und Transcodierung.....	6
Echounterdrückung.....	6
Antischall	7
Limiter, Kompressoren, Equalizer	7
Sicherung des Streams	8

Schall, den wir hören, ist einfach nur eine Variation von Luftdruck. Schall, den wir elektrisch übertragen wollen, wird anhand von Spannungs- oder Stromvariationen dargestellt. Eine digitale Darstellung kann aber nur definierte Stufen nutzen, kein analoges Ebenbild. So gesehen wird der Schall anhand von Zahlenwerten dargestellt.

Wie der Schall tatsächlich dargestellt, also codiert und decodiert (deshalb Codec) wird, hängt von den Anforderungen ab, die von der jeweiligen Anwendung vorgegeben werden. Die verschiedenen Codecs haben ihre Stärken und Schwächen, die sich aus den Parametern Latenz, Dynamik sowie Daten- und Audiobandbreite ergeben.

Nicht nur im Netzwerk, auch auf anderen Medien wie z.B. ISDN, werden solche Codecs verwendet. Die zur Verfügung stehende Datenbandbreite ist also nicht immer so hoch, wie man das in modernen Netzwerken gewohnt ist. Deshalb, und weil man oft auch mehrere Kanäle gleichzeitig betreiben möchte, sind Codecs, die zunächst schlechter erscheinen, vielfach die bessere Wahl.

Durch Codecs digital dargestellte Audiosignale, können entsprechend auch digital bearbeitet werden. Dazu folgen weiter unten auch noch gesonderte Informationen.

Codecs

Wir beschreiben hier einige sehr populäre Codecs und deren Eigenheiten. Die hier aufgezählten können wir anbieten, was wir aber nicht in allen Fällen empfehlen.

PCM

Die Pulse Code Modulation ist die wahrscheinlich einfachste Form der Codierung. Die Audiobandbreite wird durch die Abtastrate festgelegt (immer etwas mehr als die doppelte Nutzfrequenz). Ein Aliasing-Filter unterdrückt die Trägerfrequenz, die man sonst durchhören könnte.

PCM ist das, was man von den meisten WAV-Dateien auf PCs kennt. Eine CD ist z.B. mit 44,1KHz codiert, Filme mit 48KHz. In der Studioteknik verwendet man noch deutlich höhere Taktraten. Man kann aber auch nach unten skalieren. Übliche Taktraten sind dann noch 32, 24, 22,05KHz und 16KHz.

Wichtig ist auch die Auflösung des Signals in der Vertikalen. Von der CD kennen wir 16bit, aber es gibt auch feiner oder grober aufgelöste. Die Parameter ermöglichen ganz einfach die Berechnung der notwendigen Datenbandbreite für den Codec.

Beispiel: Für eine Audiobandbreite von ca. 20KHz benötigt man 44.100 Samples pro Sekunde, multipliziert mit 16bit, also 705,6Kbps pro Kanal (bekannt vom CD-Player, der aber wegen Stereo das doppelte benötigt).

ADPCM

Adaptives differentielles PCM folgt im Prinzip den gleichen Regeln wie PCM, jedoch werden die notwendigen Stufengrößen hier durch einen komplizierten Algorithmus optimiert. Man erreicht die gleiche Wiedergabequalität bei einer geringeren Auflösung in der Vertikalen.

Tatsächlich halbiert ADPCM die notwendige Datenbandbreite bei vergleichbarer Qualität. Ein Unterschied ist kaum zu hören. Man kann also auch doppelt so viele Audiokanäle auf einem Übertragungsweg transportieren.

G.711

Die PCM-Codecs sind sehr variabel, verbrauchen aber vergleichsweise viel Datenbandbreite, weil sie eigentlich für Musik optimiert sind. Das ITU hat Codecs spezifiziert, die vor allem für Sprache optimiert sind, also für die Telefonie.

G.711 ist der absolute Standard-Telefoncodec, der von jedem Gerät unterstützt werden muss, das z.B. auch SIP nutzt. Es gibt auch Chips, die G.711 direkt realisieren, ohne dass der Prozessor das rechnen müsste.

Dieser Codec legt auch gleich die Abtastrate mit 8KHz fest. Durch entsprechende Filterung liegt der Frequenzgang bei 300-3.400Hz. Unsere gerechnete Implementation reicht bis ca. 3.800Hz hinauf.

G.711 benutzt eine spezielle Codierung, um mit 8bit zur Darstellung der Amplitude auszukommen, erreicht aber eine Dynamik von 12bit. Leider gibt es zwei unterschiedliche Arten der Codierung: alaw und μ law, wobei alaw die europäische Variante ist. Unsere Implementation unterstützt beide Codierungen, es gibt dabei keinen Unterschied in Qualität oder Effektivität.

G.722

Im ISDN als „Reportercodec“ bekannt, ist G.722 ein komplexes Verfahren, das einen stark erweiterten Frequenzgang von 50-7.000Hz ermöglicht und dass bei gleicher Datenbandbreite wie G.711. Dabei ist die Dynamik so hoch, dass man sogar Musik, leicht eingeschränkt, damit übertragen kann.

Wir haben die G.722 so realisiert, dass der Frequenzgang von 150 bis 7.700Hz reicht, weil die Übertragung der tiefen Töne zu unangenehmen Nebeneffekten führen kann und man normalerweise für die Wiedergabe dieses Codecs keine Basslautsprecher verwendet. Die hohen Frequenzen sind dagegen wertvoller für die Verständlichkeit.

G.722 ist auch bekannt unter den Namen HD-Telefonie, HD-Voice oder HD-Audio. Abwandlungen des Verfahrens ermöglichen auch Reduzierungen der Datenrate von 64 auf 56 oder 48kbps. Allerdings ergeben sich deutlich hörbare Verschlechterungen, weshalb wir nur die Variante mit den 64kbps empfehlen.

G.722.1 und G.722.2 beschreiben andere Verfahren, die nicht kompatibel sind zu G.722. Beide können wir realisieren, bieten sie aber nur auf besonderen Wunsch hin an.

G.726

Hier handelt es sich wieder um einen Codec mit einer ganzen Reihe von Variationen. Die Audiobandbreite liegt bei 50-4.000Hz, ist dem G.711 also recht ähnlich. Durch den Einsatz von ADPCM wird allerdings eine geringere Datenbandbreite benötigt, die zwischen 16 und 40kbps liegt. Wir empfehlen die Variante mit 32kbps, die eine ordentliche Sprachqualität erbringt, vergleichbar mit G.711, aber nur halb so viel Daten erzeugt.

Wir haben Lösungen, die mit G.726 zwei Sprachkanäle auf einem ISDN-B-Kanal realisieren, also 4 unabhängige auf einer ISDN-Leitung, egal ob es sich um S0, UP0 oder UK0 handelt. Solche Lösungen werden oft benötigt um auf normalen Zweidrahtverbindungen über mehrere Kilometer bis zu 4 Sprachkanäle zu realisieren.

G.726 wird in dieser Variante auch sehr häufig in DECT-Telefonen eingesetzt.

Sonstige Codecs

Es gibt noch sehr viele andere Codecs (CELP, G.729, GSM, iLBC, Speex), viele sind inzwischen obsolet (wie z.B. G.721 und G.723). Manche davon sind nur für sehr spezielle Aufgaben sinnvoll, andere so rechenintensiv, dass man dafür fast einen PC braucht. Wir können die meisten dieser Codecs ebenfalls realisieren und auch spezielle proprietäre, falls das gewünscht wird.

Latenz und MP3

Ein wichtiger Parameter in Sachen Verbindungsqualität ist die Latenz, die einen wichtigen Einfluss hat. Sie ist aber nicht nur vom Codec, sondern auch von der Art und Weise abhängig, wie die Audiopakete gesammelt und verschickt werden.

Die Wirkung von Latenz kennt man aus der Praxis von mobilen Telefongesprächen, TV-Streaming über Internet u.ä. Der Empfang eines TV-Programms zeigt eine sehr spürbare Verzögerung gegenüber dem Originalsignal. Es reicht ein Blick auf die eigene Uhr und der Vergleich mit der Uhr vor den Nachrichten (es sind viele Sekunden).

Gestreamte Signale (Video oder Audio) werden in sogenannten Paketen verschickt. In einem solchen Paket befindet sich immer eine gewisse Menge an Zeit, die erstmal vergehen muss, bevor das Paket auf die Reise geht. Auf der Empfangsseite ist es umgekehrt. Das Paket muss erstmal empfangen werden, bevor man es wiedergeben kann und diese Wiedergabe startet nicht sofort.

Leider ist es mit dem Puffern einzelner Pakete nicht getan. Diese werden in normalen Netzwerken nämlich nicht so schön regelmäßig übertragen, wie man das gern hätte. Wenn sich ein Paket verspätet, ergibt sich eine Lücke in der Wiedergabe, der Empfangspuffer läuft leer. Um das zu verhindern, werden zunächst mehrere Pakete in einem sogenannten Jitter-Buffer gesammelt. So verhindert man Störungen bei der Wiedergabe.

Um eine störungsfreie Übertragung zu gewährleisten, wählt man die Puffer möglichst groß. Damit können auch längere Ausfälle des Streams überbrückt werden. Deshalb auch die hohe Latenz von TV-Streams.

In der Sprachkommunikation kann man so lange Verzögerungen nicht tolerieren. Das kann man bei mobilen Telefongesprächen gut beobachten. Führt man ein solches Gespräch mit jemandem, der im gleichen Raum ist, merkt man um wieviel später der Ton im Hörer kommt.

Der Effekt einer hohen Latenz in einem solchen Gespräch ist, dass man auf eine Antwort immer länger warten muss, als man es von einem normalen Gespräch her gewohnt ist. Am Ende fällt man sich gegenseitig ins Wort. Das kennt jeder und Schuld ist die teilweise hohe Latenz bei Mobilgesprächen.

Die Latenz errechnet sich aus der Anzahl von gepufferten Paketen in Versand und Empfang, sowie der Zeit für die Übertragung. Zur Berechnung muss man wissen, wieviel Zeit in einem einzelnen Paket codiert ist.

Schnelle Systeme können mit kleinen Puffern klarkommen, langsamere brauchen größere und verursachen damit höhere Latenzen, nur weil sie nicht so zeitnah ihre Aufgaben erledigen können. Deshalb ist unser Betriebssystem `rtos` ein Echtzeitbetriebssystem, das extrem schnell reagiert und mit wenig Vorlauf auskommt.

MP3 und ähnliche

Viele Codecs, wie z.B. auch MP3, haben andere, besondere Aufgabenstellungen. MP3 ist z.B. recht effektiv, wenn es um die Komprimierung von Musik in hoher Qualität geht. Aber MP3 bildet selbst Pakete mit Audiodaten und diese haben keine einheitliche Länge. Der Empfänger weiß also zunächst nicht, wieviel Zeit für die Wiedergabe eines Pakets benötigt wird.

MP3, wie auch andere vergleichbare Verfahren, erzeugen deshalb eine Latenz, die für Dialoge absolut unbrauchbar sind. Dafür ist die Wiedergabe von vorgefertigter Musik in guter Qualität möglich, wenn der exakte Zeitpunkt der Wiedergabe keine große Rolle spielt.

Die Wiedergabe eines MP3-Streams auf mehreren Empfängern, ist auch nur schwer zu realisieren, was für die anderen Codecs in einem Echtzeitsystem kein Problem darstellt.

Codecvergleich

In der Tabelle findet sich ein einfacher Vergleich zwischen verschiedenen Codecs, die übrigens auch unterschiedlich viel Rechenleistung benötigen. Mit Latenz ist hier die Mindestverzögerung gemeint, die sich beim Versenden und Empfangen in Datenpaketen auf einem lokalen Netzwerk, unter Verwendung von eLLan:RTOS ergibt.

Die Tabelle geht dabei von üblichen Paketlängen aus, die auch Kompatibilität zu Fremdsystemen gewährleisten. Für spezielle Anwendungen ist es auch möglich, mit kürzeren Datenpaketen und kleineren Puffern zu arbeiten. Das verbessert die Latenz, erhöht aber den Protokolloverhead beim Datenversand und die Anforderungen an die Rechenleistung des Systems.

Codec	Samples KHz	Bandbreite		Latenz ms	Dynamik bit
		Audio Hz	Daten kbps		
PCM	16	50-7700	192	70	12
PCM	16	50-7700	256	70	16
PCM	24	50-11500	384	70	16
PCM	32	50-15400	512	70	16
ADPCM	16	50-7700	96	70	12
ADPCM	16	50-7700	128	70	16
ADPCM	24	50-11500	192	70	16
ADPCM	32	50-15400	256	70	16
G.711	8	300-3800	64	80	12
G.722	16	150-7700	64	80	14
G.726	8	150-3800	32	80	12

Aushandlung

Sobald ein System mehr als einen Codec versteht, muss dieser im Verbindungsaufbau ausgehandelt werden. SIP sieht das so vor, aber auch alle anderen Verfahren sind darauf angewiesen. SDP ist das Protokoll, das gewöhnlich dafür verwendet wird, es gibt aber auch andere Methoden.

Bei Durchsagen ist es schwieriger, denn da gibt es keinen klassischen Verbindungsaufbau. Es ist notwendig den Codec im Empfang automatisch zu erkennen und wiederzugeben, beim Senden muss man ihn einstellen können.

Für die Behandlung von Durchsagen und anderen Sonderverfahren, haben wir Algorithmen entwickelt, die eine entsprechende automatische Erkennung ermöglichen. Das gilt auch für SIP, wo nicht für alle Codecs standardisierte Codes existieren.

Unser System integriert auch Geräte von Drittherstellern außerhalb der Standards, wenn dies notwendig ist.

Speicherung und Transcodierung

Die Codecs codieren den Audiostream nicht nur für die Übertragung, sondern auch für eine eventuelle Speicherung. Man kann das gleiche digitale Abbild speichern, das man auch überträgt. Dies vereinfacht die Handhabung des Speichers natürlich. Allerdings kann es sein, dass man dann mit verschiedenen Codecs im Speicher klarkommen muss, was sehr unhandlich ist.

Durch die Codierung ist es in den meisten Fällen nicht möglich, die Audiosignale in dieser Form direkt zu bearbeiten. Noch nicht einmal mischen lassen sie sich. Man muss also in diesen Fällen auf ein gemeinsames Format zurückgreifen, in dem man am besten auch gleich speichert.

Zur weiteren Verarbeitung eignet sich eigentlich nur der PCM-Codec, der leider auch den meisten Speicherplatz benötigt. Außerdem muss man dabei die höchste Samplerate verwenden, die im System vorkommt.

Eine Transcodierung, also Wandlung zwischen zwei Codecs, kann auch in Echtzeit erfolgen. Dies ist aber sehr rechenintensiv, weil dabei zumeist auch noch Vorgänge wie Up- und Downsampling notwendig sind, um auf gleiche Abtastraten zu kommen.

Um eine der folgenden Bearbeitungsstufen zu ermöglichen, ist eine Echtzeit-Transcodierung immer notwendig, weil die damit einhergehenden Rechenvorgänge eine lineare Darstellung des Audiosignals erfordern.

Echounterdrückung

Ein häufig gefordertes Leistungsmerkmal ist die Echounterdrückung. Bei Gegensprechverbindungen sind Echos eine unangenehme Begleiterscheinung, weil die Lautsprecherwiedergabe sofort wieder ins Mikrofon wandert und der Gegenseite als zusätzliches Echo des eigenen Signals zurückgegeben wird.

Solange man einen Telefonhörer oder ein Headset verwendet, sind Echos nicht relevant. Mit einer Freisprecheinrichtung sieht die Sache aber ganz anders aus.

Chiplösungen, die auf eine feste Hardwareanordnung setzen (wie in Mobiltelefonen oder in Autos), kommen im industriellen Umfeld nicht in Frage (auch aus Kostengründen), weil sie nicht flexibel genug sind.

In unserem System haben wir eine selbst entwickelte Softwarelösung (nach G.168), die sich automatisch auf ein weites Feld von akustischen Gegebenheiten einstellt (AEC = Adaptive Echo Cancellation).

Die Tatsache, dass diese Lösung aus unserer eigenen Feder stammt, erleichtert uns die Arbeit bei komplexen Aufgabe, zum Beispiel bei Systemen mit mehreren Mikrofonen oder Lautsprechern oder mit verschiedenen Frequenzgängen.

Unser AEC unterdrückt nicht nur das direkte Echo, sondern auch zusätzliche Rückwürfe aus der näheren Umgebung (bis zu einigen Metern).

AEC mit G.722 haben wir in unserem Betriebssystem auch schon auf Kleinst-CPU's realisiert, was preiswerte einkanalige Duplex-Intercomsysteme ermöglicht.

Antischall

Antischall ist eine Methode zur Auslöschung von unerwünschten Umgebungsgeräuschen. Man benötigt zwei Audiosignale von getrennten Mikrofonen. Das primäre ist dabei in Sprechrichtung ausgerichtet, das sekundäre in der Nähe, aber vom Sprecher abgewandt.

Beide Mikrofone empfangen ungefähr denselben Anteil an Nebengeräuschen, das Sprachsignal ist aber am primären Mikrofon deutlich ausgeprägter. Ein spezieller DSP-Algorithmus rechnet nun das sekundäre Signal aus dem primären heraus. Dabei werden Laufzeitunterschiede und Phasendifferenzen automatisch ausgeglichen, um ein gutes Ergebnis zu erhalten.

Limiter, Kompressoren, Equalizer

Es gibt noch eine ganze Reihe weiterer Algorithmen, die die Audioqualität verbessern können. Oft kommen diese zum Ausgleich anderer Schwächen zum Einsatz, manchmal aber einfach zur Verbesserung der Sprachverständlichkeit.

Limiter verhindern eine Übersteuerung und Verzerrung des Nutzsymbols, indem der Pegel vor Erreichen des Maximums heruntergeregelt wird. Limiter werden bevorzugt als „Notbremse“ zur Vermeidung von Überlastungen eingesetzt. Nach einer digitalen Bearbeitung ist dies jedoch nicht mehr notwendig, weil überhöhte Signale nicht übertragen werden. Stattdessen sind Limiter am Audioeingang sehr sinnvoll, dort, wo es noch zu Übersteuerungen kommen kann.

Kompressoren dagegen dienen der Einengung der Dynamik. Schwache Signale werden angehoben, starke abgeschwächt. Ein Kompressor kann auch nach einer digitalen Übertragung noch sinnvoll sein, denn er verbessert die Sprachverständlichkeit.

Die von uns implementierten Limiter und Kompressoren sind präemptiv. Das bedeutet, dass sie das Signal analysieren bevor es codiert oder direkt nachdem es decodiert wird. Im Ergebnis führt das zu einer Wiedergabe ohne unerwünschte Nebeneffekte.

Equalizer dienen der Korrektur des Frequenzganges eines Systems. Normalerweise linearisiert man damit ungerade Verläufe, gleicht damit Frequenzgangfehler von Lautsprechern oder Mikrofonen aus. Es kann aber durchaus gewünscht sein, bestimmte Frequenzen anzuheben (Sprachverständlichkeit) oder abzusenken (Gehäuseresonanzen). Audioprozessoren realisieren wir innerhalb des Betriebssystems mit DSP-Funktionalitäten.

Sicherung des Streams

Egal, welchen Codec man auch benutzt, der Audiostream ist ungesichert und kann von Jedermann abgehört werden, wenn er Zugriff auf die betreffenden Segmente des Netzwerks hat. Es ist nicht die Aufgabe eines Codecs, Inhalte zu verschleiern oder zu verschlüsseln, auch wenn es „Codieren“ heißt.

Will man einen Audiostream sichern und für Unbefugte unlesbar machen, braucht man ein zusätzliches Verschlüsselungsverfahren. Das funktioniert mit allen Codecs, fordert aber je nach Datenbandbreite mehr oder weniger Leistung vom System.

Um verschlüsselte Audiostreams zu transportieren, empfiehlt sich ein gesichertes Übertragungsprotokoll, in diesem Fall z.B. SRTP, mit den entsprechenden Algorithmen. Wir können solche Lösungen in vielen Formen anbieten.

Für weitere Informationen stehen wir Ihnen jederzeit zur Verfügung.