

Audio-Verarbeitung im Rundfunk

Inhaltsverzeichnis

1	Einleitung	1
1.1	Kompressionsverfahren	2
1.1.1	Kompression bei Telefon-Signalen	3
2	MPEG Audio-Kompressionsverfahren	4
2.1	Eigenschaften des Gehörs	4
2.1.1	Verdeckung bei Dauersignalen	4
2.1.2	Vor- und Nach-Verdeckung	6
2.1.3	Lautheit und Tonheit	6
2.1.4	Frequenz-Gruppen	7
2.1.5	Skalen der Basilarmembran	8
2.2	MPEG-Audio-Kompression	9
2.2.1	MPEG 2 Audio-Codierung	10
3	Kaskadierung von Strecken mit MPEG-Codierung	11
4	Audioprocessing am UKW-Sender	12

Abbildungsverzeichnis

1.1	Prinzip der Rundfunk-Übertragung	1
1.2	Zeit-Diskretisierung	1
1.3	Nachrichten-Anteile	2
1.4	A-Kennlinie	3
1.5	μ - und A-Kennlinie	4
2.1	Lautstärke-Kurven	5
2.2	Verdeckungs-Kurven für Schmalband-Rauschen 1kHz	5
2.3	Verdeckungs-Kurven für Schmalband-Rauschen	5
2.4	Verdeckungs-Kurven für Violinklänge	6
2.5	Vor- und Nach-Verdeckung	6
2.6	Tonheit	7
2.7	Lautheit	7
2.8	Frequenzgruppen	7
2.9	Tonheit — Frequenzgruppe	8
2.10	Frequenz logarithmisch über Tonheit	8
2.11	Verdeckung bei einem Vokal	9
2.12	Verdeckung	9
2.13	Blockschaltbild MPEG 2 Encoder	10
2.14	Echtzeit-Analyzer und Filterbank	10
2.15	Filterbank MPEG Encoder	11
4.1	Optimod	12

Audio-Verarbeitung im Rundfunk

1 Einleitung

Zu Beginn des Rundfunks waren die Verhältnisse noch recht übersichtlich. Ein Audio-Signal wurde von einem Mikrophon in ein elektrisches Signal gewandelt und per (analoger) Rundfunk-Leitung zum Sender übertragen, wo es durch den Modulationsvorgang die Trägerwelle beeinflusste, welche dann über den Sendemast ausgestrahlt wurde, Bild 1.1. [1] Damals war die gesamte Übertragungs-Strecke noch analog.

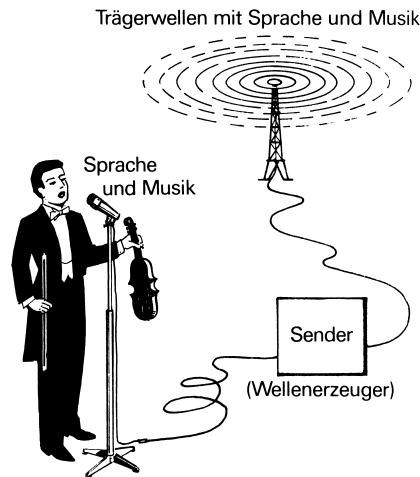


Bild 1.1: Prinzip der Audio-Übertragung in den Anfangszeiten des Rundfunks

Anfänglich kamen im Prinzip nur etwas „bessere“ Telefonleitungen, dann spezielle (fest geschaltete) Rundfunk-Leitungen zum Einsatz. Die Audio-Signale, die heute im digitalen Studio einer Rundfunkanstalt aufgenommen werden, bestehen aus PCM Daten (Puls-Code-Modulation mit 16 Bit bzw. 24 Bit).

Zeit-diskrete Signale Der erste Schritt zur Digitalisierung eines Audio-Signals ist seine zeitliche Diskretisierung, indem zu äquidistanten Abtast-Zeitpunkten Stützwerte aus dem analogen Verlauf entnommen werden, Bild 1.2.[3]

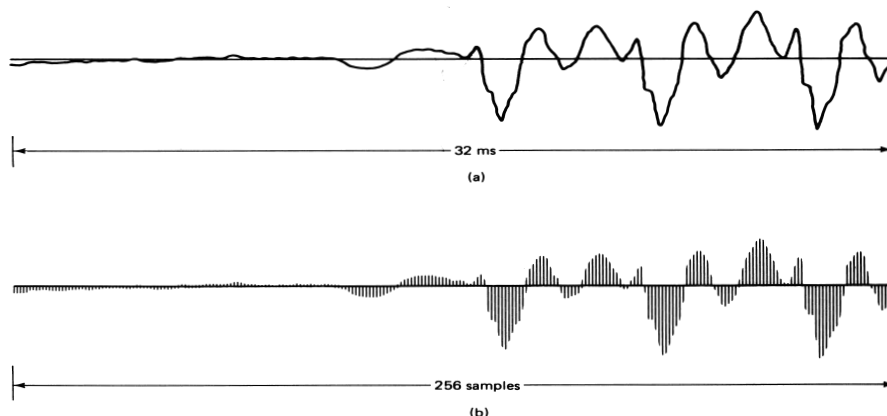


Bild 1.2: Ausschnitt aus einem Zeit- und Amplituden-kontinuierlichen Sprachsignal (a) und Zeit-diskrete, aber Amplituden-kontinuierliche (analoge) Abtastwerte mit $T_A = 125 \mu\text{sec}$ entsprechend einer Abtastfrequenz $f_A = 8 \text{ KHz}$ (Beispiel: Telefon-Signal, Grenzfrequenz $f_g = 3,4 \text{ KHz}$)

Die Amplituden-Werte der Stichproben (*Samples*) in Bild 1.2 (b) sind noch nicht quantisiert (bzw. digi-

talisiert), weswegen es sich hier um ein Zeit–diskretes analoges Signal handelt.*¹ Ein derartiges Zeitsignal wird auch als Puls–Amplituden–Modulation (PAM) bezeichnet.*²

Gemäß einem Theorem von Shannon kann der Zeit–kontinuierliche Signalverlauf, Bild 1.2 (a), durch Tiefpaß–Filterung exakt aus dem Zeit–diskreten Signalverlauf, Bild 1.2 (b), zurückgewonnen werden.[10]

Im nächsten Schritt müssen die Amplituden–Werte der Samples quantisiert werden. Entsprechend zu den gewählten Amplituden–Stufen können die quantisierten Samples dann mit Hilfe von Bits und Bytes dargestellt werden. Entsprechende A/D–Wandler finden sich heute in jedem PC, ebenso die zugehörigen D/A–Rückwandler. Die Auflösung ist in diesem Falle 16 Bit (32768 Amplitudenstufen linear) bei 44,1 KHz Abtastrate. Das ergibt eine Datenrate von $44100 \cdot 16 \cdot 2$ bit/sec (1,4112 Mbit/sec) für ein Stereo–Signal.*³

Speicherung Da die Programme i.a. nicht „live“ gesendet werden, müssen diese zuvor gespeichert werden. Die Ton–Archive beim Rundfunk sind heute (fast) nur noch digital. Die Kosten für die (digitale) Speicherung sind ebenfalls direkt von der Datenrate bzw. Datenmenge abhängig. Somit gelten für die Speicherung bzw. für das digitale Audio–Archiv des Rundfunks gleichartige Überlegungen wie für die Übertragung digitalisierter Audiosignale. (MPEG1 Layer2 mit 384 kBit/sec.)

Zubringer zum Sender Die Modulations-Inhalte müssen über „Zubringer–Leitungen“ zu den Sendern übertragen werden. Die Art der Übertragung kann (auch noch) analog oder (meist) digital geschehen und diese „Leitungen“ sind z.B.: Glasfasern (optische Leitungen, digital), Funkstrecken (terrestrisch, analog oder digital), Satelliten (digital) oder Kupferleitungen (analog oder digital).

Die Kosten für diese Leitungen richten sich in jedem Fall nach der erforderlichen Bandbreite (bzw. der Datenrate) für die Übertragung, da alle „Leitungen“ mehrfach belegt werden können. Je geringer die notwendige Bandbreite ist, um so mehr Teilnehmer (Kunden) können pro Leitung bedient werden. Die Kosten für die Übertragung werden minimiert, wenn möglichst viel Bandbreite gespart wird. Daraus resultiert die Forderung nach geringen Datenraten bei digitaler Übertragung. (384kBit/sec, 256 kBit/sec, 192 kBit/sec)

Ein wichtiges Kriterium beim Rundfunk ist die Qualität (Frequenzumfang, Dynamik, Klirren und Verzerrungen, Geräuschabstand) einer Übertragung. Die Anforderungen an die Qualität begrenzen schließlich auch den Faktor, um den ein Audiosignal zwecks Einsparung von Bandbreite komprimiert werden kann.

1.1 Kompressionsverfahren

Redundanz und Relevanz Bei den Kompressionsverfahren allgemein unterscheidet man zwischen solchen, die nicht verlustbehaftet sind und die redundante Teile entfernen, und verlustbehafteten, die zusätzlich die irrelevanten Teile entfernen, Bild 1.3.

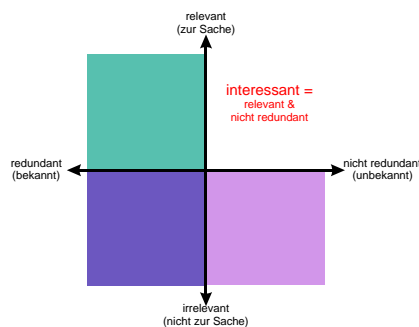


Bild 1.3: Eine Nachricht allgemein, bzw. ein Audio–Signal speziell, setzt sich stets aus mehreren Anteilen zusammen. Zur Übertragung wird nur der relevante und nicht redundante Anteil benötigt. Verlustlose Kompressionsverfahren beseitigen nur den redundanten Anteil. Verlustbehaftete Verfahren beseitigen auch den irrelevanten Anteil.

*¹In der Literatur wird oftmals in der Darstellung nicht zwischen Zeit–diskreten (analogen) und Zeit–diskreten, Amplituden–quantisierten (digitalen), Signalen unterschieden.

*²„Modulation“ bedeutet bei Puls–Modulationen keine Frequenz–Verschiebung im Spektrum wie diese bei Modulationen mit cosinusförmigen Trägersignalen erfolgt.

*³Im Falle von Telefon–Signalen, Bild 1.2, werden 8 Bit (1 Byte) verwendet, was auf ± 128 Amplitudenstufen führt. Der daraus resultierende Datenstrom wird somit $8000 \cdot 8$ bit/sec (64 Kbit/sec), was einem ISDN–Kanal entspricht.

Bei nicht verlustbehaftete Verfahren können die originalen Daten wieder rekonstruiert werden (Beispiel .zip), während bei verlustbehafteten Verfahren größere Kompressionsfaktoren erreichbar sind, jedoch die originalen Daten nicht wieder hergestellt werden können. Wichtig ist dann nur, daß der Anwender (z.B. das Ohr oder das Auge) davon praktisch nichts merkt.

Beispiel eines verlustbehafteten Verfahrens Ein klassisches verlustbehaftetes analoges (!) Verfahren ist die Wiedergabe des Fernsehbildes. Auf dem Bildschirm einer Fernsehröhre z.B. flitzt ein „wild gewordener“ Leuchtpunkt zeilenweise von oben nach unten. Trotzdem kann man den Eindruck gewinnen, „alles“ gesehen zu haben. Beim Auge sind wir gewohnt, deartig „betrogen“ zu werden. Man genießt auch einen Film im Kino und denkt nicht darüber nach, daß man eigentlich mehr als die Hälfte der Zeit im Dunkeln gesessen hat, weil ja die einzelnen Bilder immer nur kurz aufblitzen. Aber die Trägheit des Auges interpoliert die einzelnen Lichtblitze zu einem kontinuierlichen bewegten Bild.

Mit analoger Technik war es nicht notwendig und auch nicht möglich, ein Audio-Signal so zu komprimieren, daß es einen geringeren Bedarf an Speicherplatz hatte.*⁴ Mit der Digitalisierung der Audiotechnik ist ein Bedarf für eine Kompression entstanden, um die Datenraten zu verkleinern.

1.1.1 Kompression bei Telefon-Signalen

Bereits beim Telefon wurde ein Kompressionsverfahren angewendet, um die Datenrate zu begrenzen. Die Grundidee für dieses Kompressionsverfahren beruht auf einer Eigenschaft des Ohres, nämlich der, daß bei lauten akustischen Signalen gleichzeitige leise Signale nicht wahrgenommen werden. In der Akustik ist das als „Simultan-Verdeckung“ bekannt. Bei der beim Telefon angewendeten Puls-Code-Modulation (PCM) wird dieser Verdeckungs-Effekt dazu genutzt, daß laute Signale grober quantisiert werden können als leise Signale. Das Quantisierungs-Geräusch, das bei groberer Quantisierung (geringere Bit-Auflösung) steigt, wird vom lauten Signalpegel verdeckt und bleibt unhörbar. Nach der oben gegebenen Definition werden nur teilweise die irrelevanten Teile der Nachricht eliminiert. Es handelt sich also um ein verlustbehaftetes Verfahren.

Als Vorschrift für die pegelabhängige Quantisierung gilt in Europa die A-Kennlinie, Bild 1.4.[4]

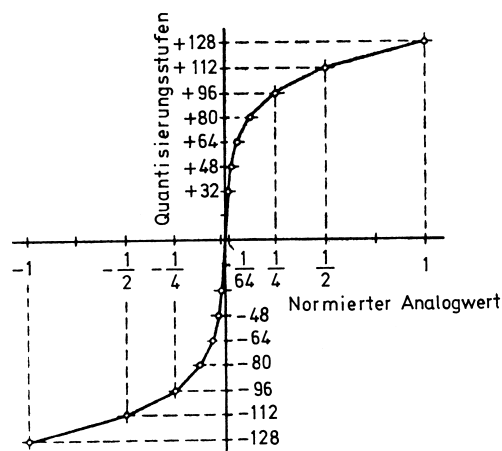


Bild 1.4: Die A-Kennlinie zur Quantisierung für die PCM beim Telefon

Wie man leicht ausrechnen kann, wären bei einer gleichmäßigen Quantisierung (mit der kleinsten Stufenhöhe) 4096 Stufen – entsprechend zu 12 Bit – statt der 256 Stufen (8 Bit) erforderlich. Die erzielte Ersparnis von 64 Kbit/sec gegenüber 96 Kbit/sec ist bei diesem Kompressionsverfahren zwar nicht sehr markant, jedoch für das Telefonsystem trotzdem wesentlich.

In den USA wird die μ -Kennlinie verwendet, die etwas anders verläuft, wie der Vergleich in Bild 1.5 verdeutlicht.[5]

Obwohl also die Kompressions-Kennlinien nicht exakt übereinstimmen, wird das z.B. bei einem transatlantischen Telefongespräch nicht bemerkt.

*⁴Rauschunterdrückungsverfahren wie Dolby, HighCom oder Optimod verändern hierbei nur die Dynamik des Audiosignals.

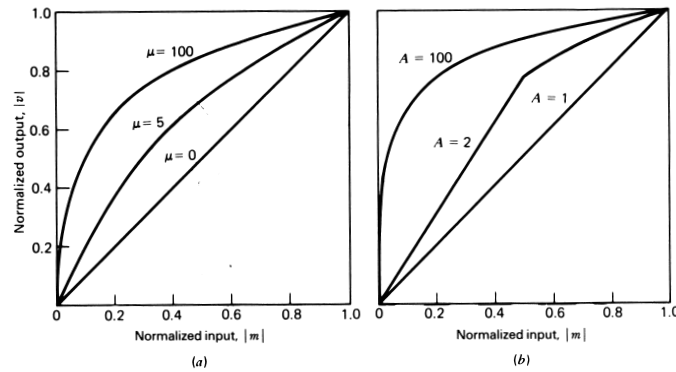


Bild 1.5: Vergleich von μ - und A-Kennlinie zur Quantisierung für die PCM beim Telefon

2 MPEG Audio-Kompressionsverfahren

Für digitale Audio-Übertragungen im Rundfunk werden Kompressionsverfahren nach MPEG (*moving pictures expert group*) eingesetzt. Für Satelliten und Zubringer für terrestrische UKW-Sender, aber auch für AM-Sender im LMK Bereich kommt MPEG1-Layer2 (kurz: MPEG2) zum Einsatz, das nachfolgend beschrieben werden soll.

Bei den MPEG Audio-Kompressionsverfahren handelt es sich um verlustbehaftete Verfahren, die nicht nur die Redundanz, sondern auch die Irrelevanz aus den Audiosignalen entfernen. Sie sind international genormt, da sie allgemein Anwendung finden. Bei der Beschreibung für international genormte Verfahren wird stets das Signal beschrieben und zusätzlich entweder die Sender- oder die Empfänger-Seite, je nachdem was einfacher ist. Im Falle der Audio-Kompressionsverfahren ist die Expander-Seite (Empfänger) definiert, da diese einfacher ist.

Wie der Kompander zu realisieren ist, wird somit der Ingenieurskunst der jeweiligen Herstellerfirma überlassen. Das führt aber dazu, daß nicht alle Geräte sich völlig gleichartig verhalten. Man kann aber davon ausgehen, daß ein Signal, das den Kompander und den Expander einer Firma durchlaufen hat, sich innerhalb der von der Norm vorgegebenen Grenzen hält.

Wenn jedoch auf einer Zubringerstrecke mehrere MPEG Kompressions-Strecken von unterschiedlichen Firmen kaskadiert werden, können jedoch — programmabhängig — wahrnehmbare Unterschiede entstehen.[2]

2.1 Eigenschaften des Gehörs

Alle (verlustbehafteten) Audio-Kompressionsverfahren nutzen die Eigenschaften des Gehörs aus. Die Idee dabei ist, daß alles das weggelassen werden kann, bzw. nicht übertragen werden muß, was das „normale“ Ohr nicht hört. Informationstechnisch bedeutet das, daß aus dem Audio-Signal alle (für das Ohr) redundanten und irrelevanten Anteile entfernt werden. Diesbezügliche Untersuchungen gab es schon lange vor einer Anwendung für eine Audiokompression.[7] [8]

Das Ohr hat eine frequenzabhängige Hörschwelle und wie der Frequenzgang sich bei unterschiedlichen Lautstärken verhält, ist schon seit Längerem bekannt, Bild 2.1.[6] Die Hörfähigkeit verschlechtert sich mit zunehmendem Alter, was besonders die hohen Töne betrifft.

2.1.1 Verdeckung bei Dauersignalen

Die Hörschwelle ändert sich jedoch auch bei Anwesenheit eines Tones oder Geräusches. Das bewirkt, daß durch diesen Ton bzw. dieses Geräusch ein anderes akustisches Signal nicht wahrgenommen werden kann und somit „verdeckt“ wird. Das wurde bereits, wie oben beschrieben, bei der 13-Segment-Kennlinie zur Gewinnung des PCM Signals beim Telefon angewendet.

Die akustische Forschung hat sich jedoch schon vor einer Anwendung in der Technik mit dem Effekt der Verdeckung (*masking*) beschäftigt.^{†1} Bild 2.2 [8] zeigt die Verdeckungs-Kurven oder „Mithör-Schwellen“

^{†1}Beim Studium zu Beginn der '60er Jahre galt die Wahl-Vorlesung „Das Ohr als Nachrichtenempfänger“ an der UNI Stuttgart eigentlich noch als „Exoten-Fach“. Die Studenten der Vorlesung waren z.T. auch die „Versuchskaninchen“ bei den Meßreihen, in denen die Erkenntnisse zu den Eigenschaften des Ohres gewonnen wurden.

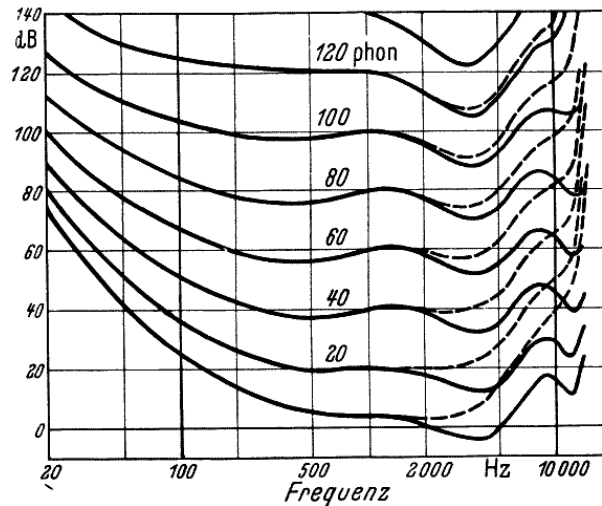


Bild 2.1: Hörschwellenkurve und Kurven gleicher Lautstärke (— : im Alter von 20 Jahren; - - - : im Alter von 60 Jahren)

für ein Schmalband-Rauschen der Mittenfrequenz 1 kHz und unterschiedlicher Lautstärke. Verdeckungs-Kurven für Schmalband-Rauschen mit gleicher Lautstärke und unterschiedlichen Mittenfrequenzen sind in Bild 2.3 [8] dargestellt.

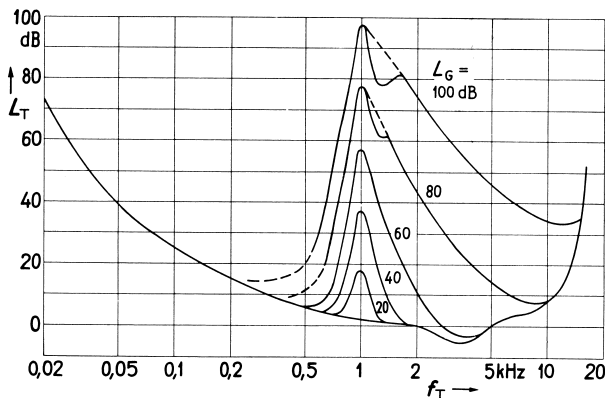


Bild 2.2: Verdeckungs-Kurven für Schmalband-Rauschen bei 1 kHz unterschiedlicher Lautstärke

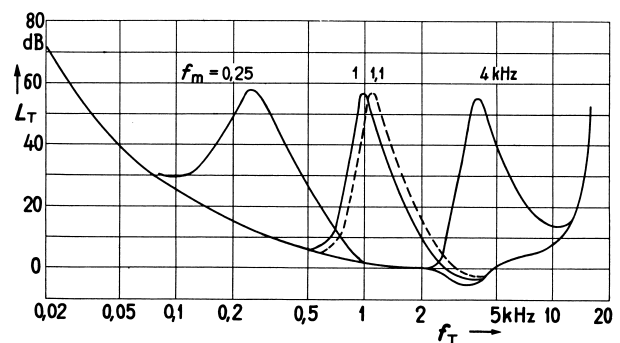


Bild 2.3: Verdeckungs-Kurven für Schmalband-Rauschen unterschiedlicher Mittenfrequenz

Alle akustischen Signale, deren Lautstärke unterhalb der Verdeckungs-Kurven liegt, werden nicht wahrgenommen. Man muß sie deshalb auch nicht übertragen. Eine Möglichkeit, diese Eigenschaft zu nutzen, gab es bei analoger Übertragung nicht. Erst nachdem eine digitale Verarbeitung möglich ist — und damit eine schnelle Fourier-Transformation (FFT *fast fourier transform*) — kann dies realisiert werden.

Die Verdeckungs-Kurven sind nur empirisch zu ermitteln und haben einen komplizierten Verlauf, Bild 2.4.[8] Sie können daher nur schwer in Formeln gegossen werden. Aber genau das wird (eigentlich) benötigt für einen MPEG Audio-Encoder. Jeder Hersteller von MPEG Audio-Encodern verwendet hier seinen eigenen Algorithmus zur Nachbildung der Verdeckungskurven, der als „Firmengeheimnis“ gilt. So ergeben sich zwangsläufig gewisse Unterschiede zwischen den Encodern der verschiedenen Anbieter.

Dieses Beispiel, Bild 2.4, zeigt eine allgemeine Eigenschaft des Ohrs:

- Leise hohe Töne werden durch tiefe laute Töne verdeckt, Bild 2.4 (links).
- Leise tiefe Töne werden durch hohe laute Töne nicht verdeckt, Bild 2.4 (rechts).^{†2}

^{†2}In einem gemischten Chor sind deshalb Männerstimmen immer schwächer besetzt als Frauenstimmen. In einem Streichorchester ist die Zahl der Celli nur ein Bruchteil der Zahl der Violinen.[8]

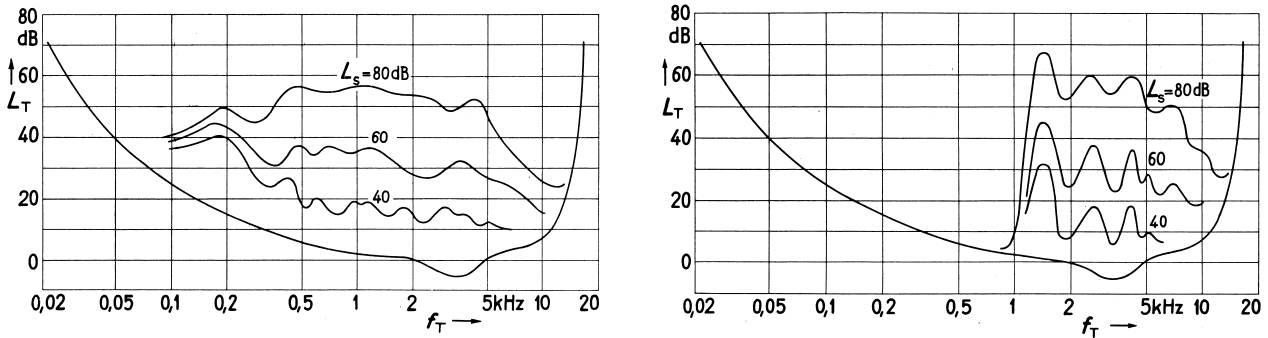


Bild 2.4: Verdeckungs-Kurven für tiefe und hohe Violinklänge

Die technische Anwendung des Verdeckungs-Effektes erfordert eine Analyse des Schall-Ereignisses im Frequenzbereich. Hierzu ist entweder geeignete Filter-Bank zu verwenden (MPEG1 & MPEG2) oder das Schall-Signal muß in den Frequenzbereich transformiert werden (MPEG3). Beide Ansätze werden technisch ausgeführt und führen auf unterschiedliche Kompressions-Verfahren.

2.1.2 Vor- und Nach-Verdeckung

Eine Verdeckung gibt es aber nicht nur bei Anwesenheit eines Dauer-Tones bzw. Dauer-Geräusches. Auch das, was kurz vor bzw. kurz nach einem lauten Geräusch als akustisches Signal auf das Ohr trifft, wird verdeckt, Bild 2.5.[9]

Diese Eigenschaft des Ohrs ist für die Anwendung des Audio-Kompressionsverfahrens deshalb von Bedeutung, weil zur Transformation des Schall-Signals in den Frequenzbereich dieses in kleine Abschnitte aufgespalten werden muß um eine FFT bzw. eine MCT (*modified cosine transform*) anwenden zu können.^{†3}

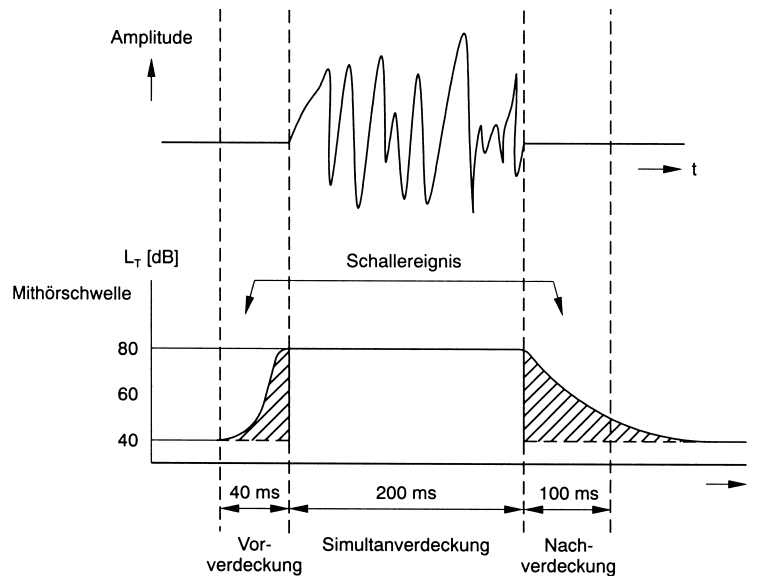


Bild 2.5: Signale, die kurz vor bzw. kurz nach einem Schallereignis auf das Ohr treffen, werden verdeckt: Vor- bzw. Nach-Verdeckung

2.1.3 Lautheit und Tonheit

Zwischen dem physikalischen Reiz (Lautstärke bzw. Tonhöhe) und der subjektiven Empfindung (Lautheit bzw. Tonheit) besteht kein linearer (proportionaler) Zusammenhang. Untersuchungen über die Empfindung

^{†3}Dies wird bei MPEG3 angewendet.

von Lautstärke, die „Lautheit“ und die Empfindung von Tonhöhen, die „Tonheit“^{†4}, zeigen die Bilder 2.7 und 2.6.[7] [8]

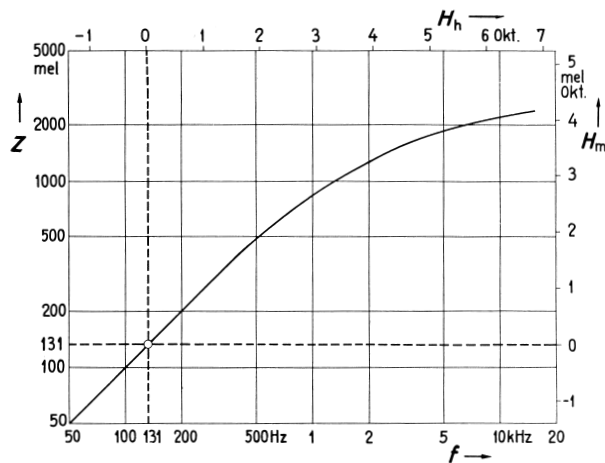


Bild 2.6: Zusammenhang zwischen der Frequenz und der empfundenen Tonhöhe (Tonheit z ; „melodische“ Oktaven) eines Tones. Der Ton c_0 mit 131 Hz ist Bezugspunkt.

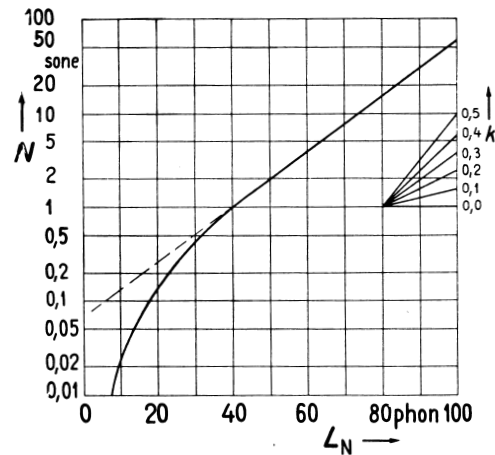


Bild 2.7: Zusammenhang zwischen Lautstärkepegel L_N und der empfundenen Lautstärke (Lautheit N) eines Tones.

Bei der (melodisch) empfundenen Tonhöhe (Tonheit z , Einheit „mel“) ist bis ca. 500 Hz Proportionalität zur physikalischen Frequenz f (Hz) gegeben. Bei höheren Frequenzen existiert keine Proportionalität. Das hat zur Folge, daß physikalisch ca. 7 Oktaven bis 20 KHz bestehen, „melodisch“ jedoch nur 4,3 (melodische) Oktaven bis zu empfundenen ca. 2300 mel.^{†5}

2.1.4 Frequenz-Gruppen

Das Gehör kann einerseits mehr als 600 verschiedene Tonhöhen unterscheiden, faßt aber andererseits bestimmte Frequenzgebiete zu „Frequenz-Gruppen“ zusammen, Bild 2.8.[8]

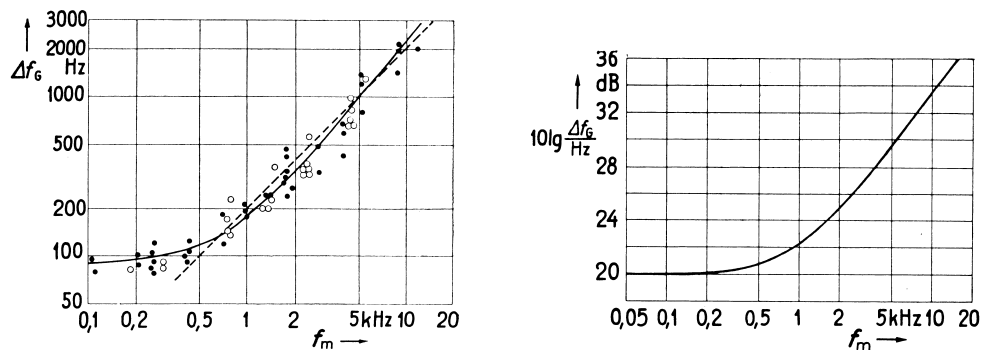


Bild 2.8: Breite Δf_G der Frequenz-Gruppen als Funktion der Mitten-Frequenz f_m ; mit Meßpunkten (links), als Pegel (rechts)

Das bedeutet, daß die Schalleistungen von (gleich lauten) Tönen verschiedener Frequenzen innerhalb einer Frequenz-Gruppe sich addieren. Die Breite Δf_G der Gruppen ist bis ca. 500 Hz konstant ≈ 100 Hz und steigt bei höheren Frequenzen proportional zur Mitten-Frequenz f_m der Gruppen bis auf 3,5 KHz an.

Das Gehör kann eine Frequenzgruppe an jeder beliebigen Stelle der Frequenzskala bilden, d.h. die Mitten-Frequenz kann beliebig im Hörbereich liegen. Teilt man die Frequenzachse, beginnend bei 20 Hz,

^{†4}Harmonische Obertöne hoher Ordnung klingen deswegen dissonant.

^{†5}In einer Testreihe sollte die Frequenz gefunden werden, die halb so hoch wie ein 8 KHz Schmalbandrauschen empfunden wird. Die Messung ergab, daß eine Frequenz von ca. 1,35 KHz als halb so hoch wie empfunden wurde wie das 8 KHz Schmalbandrauschen. [8]

bis 16 KHz lückenlos in Frequenz-Gruppen ein, so findet man, daß 16 Frequenzgruppen hineinpassen. Hier hat man dann feste Mitten-Frequenzen.^{†6}

Werden die 16 Frequenz-Gruppen in die Kurve für die Tonheit, Bild 2.6, eingetragen, so findet man, daß die Meßpunkte genau auf der Kurve liegen, Bild 2.9. Das Ohr empfindet offensichtlich jede (unterschiedlich breite) Frequenz-Gruppe als „Ton-Gruppe“ der (gleichen) Breite 100 mel. Diese Breite „100 mel“ der Tongruppen wird mit „1 Bark“^{†7} bezeichnet.

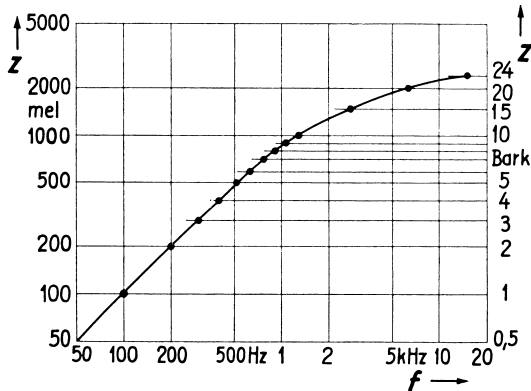


Bild 2.9: Tonheit z und Frequenz-Gruppenskala als Funktion der Frequenz; Die Meßpunkte liegen exakt auf der Kurve.

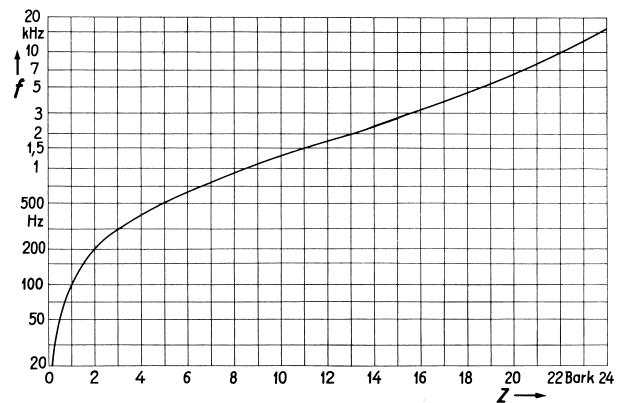


Bild 2.10: Frequenz (logarithmisch) über Tonheit z , beziffert in Bark (1 Bark = 100 mel)

Wird die Frequenz f logarithmisch über der Tonheit z , beziffert in 100 mel = 1 Bark, aufgetragen, ist erkennbar, daß oberhalb von 500 Hz ein logarithmischer Zusammenhang zwischen Tonheit und Frequenz besteht, Bild 2.10.

2.1.5 Skalen der Basilarmembran

Die Eigenschaften des Ohrs bezüglich Empfinden der Höhe einer Frequenz, also der Tonheit, ist in den anatomischen Eigenschaften der Basilarmembran zu finden, auf der die Sinneszellen sitzen (das Corti'sche Organ), die die zugehörigen Signale an das Gehirn weiterleiten. Die Länge der Membran ist 32 mm. In 4 Reihen sitzen je 3500 Haarzellen, deren gegenseitiger Abstand somit $9 \mu\text{m}$ beträgt. Sie sind über die gesamte Länge der Basilarmembran gleichmäßig verteilt.

Das Ohr kann 620 Tonhöhenstufen unterscheiden. Bezogen auf die Basilarmembran hat eine Tonhöhenstufe eine Länge von $52 \mu\text{m}$, an der dann ca. 5,5 Haarzellen (in jeder der 4 Reihen) beteiligt sind. Unterhalb von 500 Hz findet man davon 140 Tonhöhenstufen, die da identisch mit Frequenzstufen sind. Oberhalb von 500 Hz bis 16 KHz sind es noch 480 (gleich breite) Tonhöhen-Stufen, entsprechend zu (breiter werdenden) Frequenz-Stufen. An jeder Stufe sind aber gleich viele Haarzellen beteiligt.

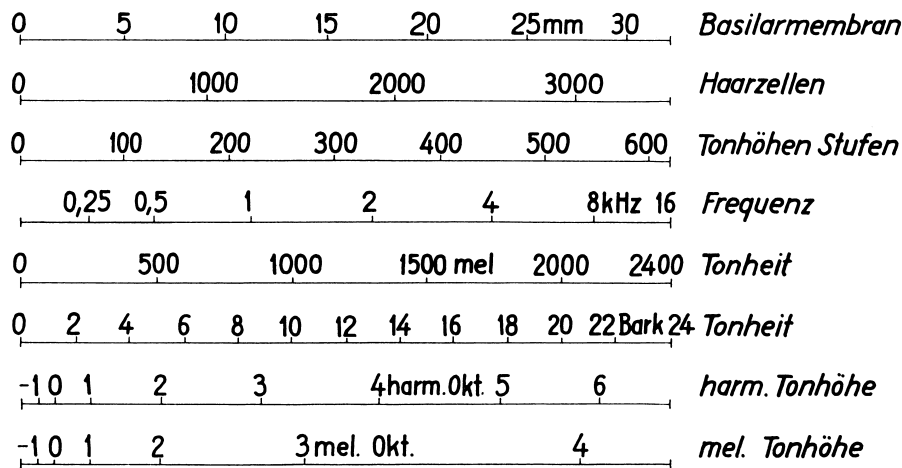
Die Skala der Tonhöhenempfindung (Tonheit) reicht bis 2400 mel und ist bezogen auf die Basilarmembran linear unterteilt. Damit ist jede der 620 Tonhöhenstufen 3,9 mel. Wegen der linearen Unterteilung in mel ist die Basilarmembran auch bezüglich der Ton-Gruppen (1 Bark = 100 mel) linear unterteilt. Die Breite einer Ton-Gruppe (und damit auch einer Frequenz-Gruppe) auf der Basilarmembran ist somit $32\text{mm}/24 = 1,3 \text{ mm}$, was 150 Haarzellen entspricht.

Das Ohr empfindet das akustische Signal einer Frequenz-Gruppe als Ton-Gruppe, das einen Abschnitt von 150 Haarzellen (in jeder der 4 Reihen) anregt. Bis zu Frequenzen von ca. 500 Hz stimmt das Resonanzverhalten der Basilarmembran mit der Aufteilung der Haarzellen überein, so daß hier Proportionalität besteht (1 mel = 1 Hz). Zwischen 500 Hz und 16 KHz besteht diese Proportionalität dagegen nicht mehr.

In der Tabelle sind die Skalen der Basilarmembran zusammengestellt. [8]

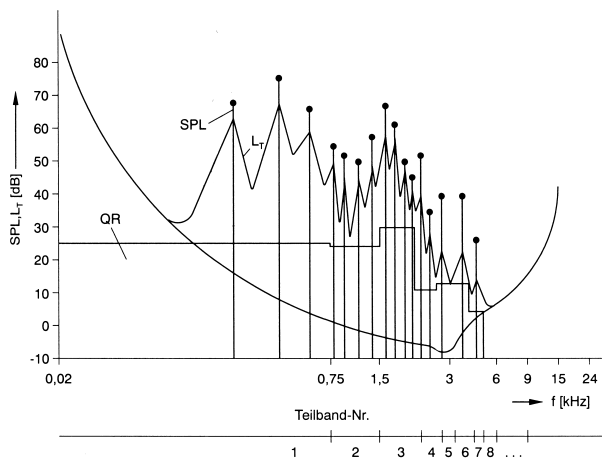
^{†6}Eine Filterbank zur spektralen Analyse des Schalls hat ebenfalls feste Mitten-Frequenzen.

^{†7}„Bark“ in Erinnerung an Heinrich Barkhausen, der sich als erster mit der Messung der Lautstärkeempfindung beschäftigte. [8]



2.2 MPEG–Audio–Kompression

Bild 2.11 zeigt die Spektralverteilung der Amplituden für einen Vokal und die daraus resultierende Mithörschwelle (Verdeckungs–Schwelle).^{†8}



SPL: Amplitudenspektrum (typ. Vokal) mit 16 Teilbändern
 Lt: Mithörschwelle
 QR: Quantisierungsrauschen

Bild 2.11: Spektralverteilung eines Vokals und die zugehörige (prinzipielle) Verdeckungskurve. Die erforderliche Bit–Auflösung richtet sich danach, wie weit das Quantisierungs–Geräusch *QR* verdeckt wird.

Die Bit–Auflösung kann so klein gewählt werden, daß das Quantisierungs–Geräusch gerade noch verdeckt wird. Das ist zunächst eine ähnliche Überlegung wie schon bei der 13–Segment–Kennlinie beim Telefon.

Bei der MPEG–Audio–Codierung kommt nun noch hinzu, daß alle die Spektralkomponenten nicht übertragen werden, die sich unterhalb der Verdeckungskurve eines laueren Tones befinden, Bild 2.12.[11] Es ergibt sich so eine wesentliche Reduktion des digitalen Datenstromes.

Die spektrale Zerlegung eines Audio–Signals muß dabei nicht mit beliebig schmalen Filtern erfolgen, vielmehr genügt es, die Analysebreite entsprechend zu der Breite der Frequenz–Gruppen zu machen.

^{†8}Da die Art und Weise der Kompression — im Unterschied zur Expansion — nicht beschreiben ist, kann demzufolge auch nur die prinzipielle Funktionsweise der Kompression angegeben werden.

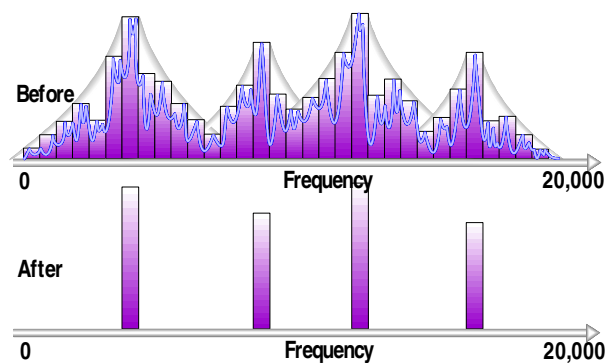


Bild 2.12: Spektralverteilung eines akustischen Signals, bestimmt aus den Ausgangssignalen einer Filterbank mit 32 gleich breiten Analyse–Filtern. Es sind Verdeckungs–Kurven (grau) gezeichnet, durch die Teil–Töne verdeckt werden. Diese müssen nicht übertragen werden.

2.2.1 MPEG 2 Audio-Codierung

Das Blockschaltbild des MPEG 2 (MPEG1–Layer2) Encoders zeigt Bild 2.13.[12]

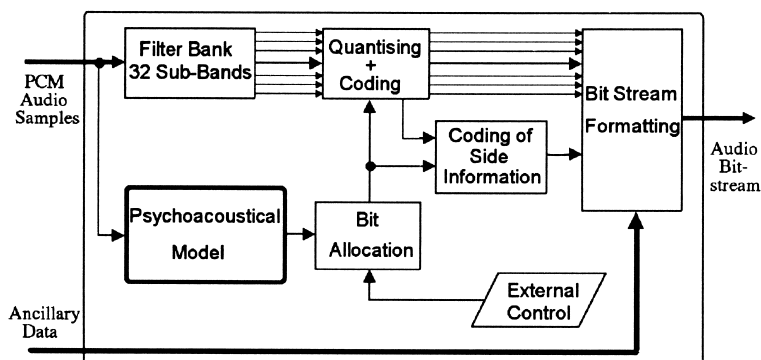


Bild 2.13: Blockschaltbild des MPEG1–Layer2–Encoders (MPEG 2)

Das Eingangs–Signal ist ein digitales PCM–Signal (gleichförmige Quantisierung mit 16 Bit), wie es z.B. von einem A/D–Wandler oder einer Audio–CD geliefert wird. Dieses Signal wird mit Hilfe einer digitalen Filter–Bank in 32 Teil–Bänder (*sub bands*) zerlegt. Dies entspricht einer (näherungsweise) spektralen Analyse in Echtzeit, Bild 2.14. [13]

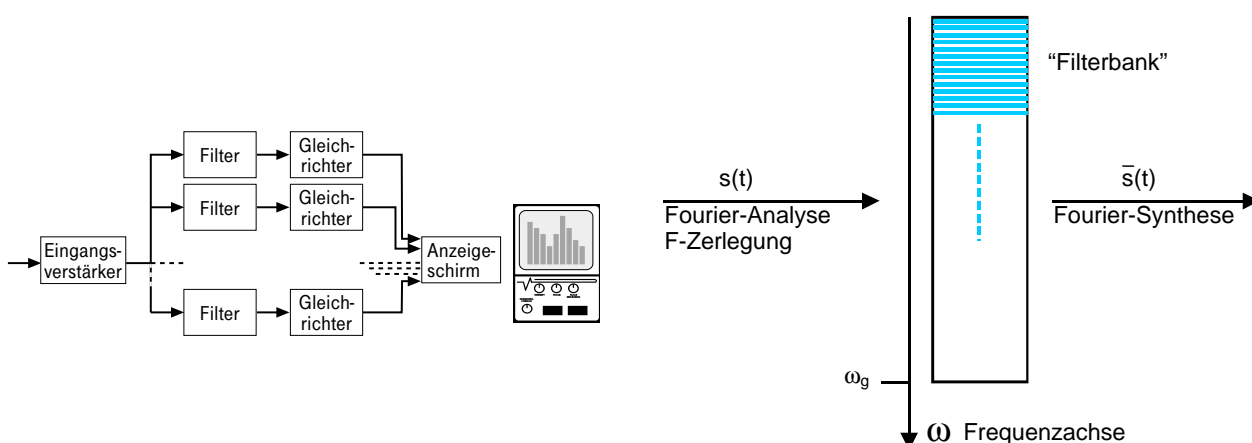


Bild 2.14: Echtzeit–Frequenz–Analysator (links); Filterbank als (physikalisches) Modell für eine Fourier–Analyse bzw. Fourier–Synthese (rechts)

Bei MPEG Encodern besteht die Filterbank aus 32 Teilband–Filtern (*subband filter*) der Breite $B_s = 625$ Hz, wodurch sich eine Analyse–Bandbreite von 20 KHz ergibt, Bild 2.15 (s. auch Bild 2.12).^{†9} Die Abtastwerte (*samples*) [je 2 Byte] aus dem Eingangssignal erzeugen entsprechende Samples in den Subband–Filtern.

Jeweils 12 Samples aus jedem der Subband–Filter werden zu einem Rahmen (*frame*) bei MPEG1 (Layer 1) zusammengefaßt, während bei MPEG2 und MPEG3 (Layer 2, Layer 3) jeweils 36 Samples aus jedem Subband–Filter zu einem Rahmen zusammengefaßt werden.

Damit die Filterung „in Echtzeit“ durchgeführt werden kann, haben die digitalen Subband–Filter nur relativ wenige Koeffizienten, was eine geringe Flankensteilheit dieser Filter zur Folge hat. Töne mit einer Frequenz an der Bandgrenze eines Filters werden daher auch im anschließenden Filter (schwächer) erscheinen.

Das „Psychoakustische Modell“ ist Firmengeheimnis (und daher nicht im Detail bekannt) und der schwierigste Teil des Encoders, denn hier sind die Eigenschaften des Ohrs zu berücksichtigen.^{†10} In diesem Block werden die Verdeckungskurven berechnet, die einerseits bestimmen, mit wieviel Bit jeweils codiert werden

^{†9}Zur Breite der Subband–Filter findet sich auch die Angabe $B_s = 750$ Hz, was dann eine Analyse–Bandbreite von 24 KHz ergibt.

^{†10}Diese wiederum können nur in langen und aufwändigen Meßreihen mit ausreichend vielen Teilnehmern ermittelt werden und stellen demzufolge irgend welche gemittelten Werte dar.

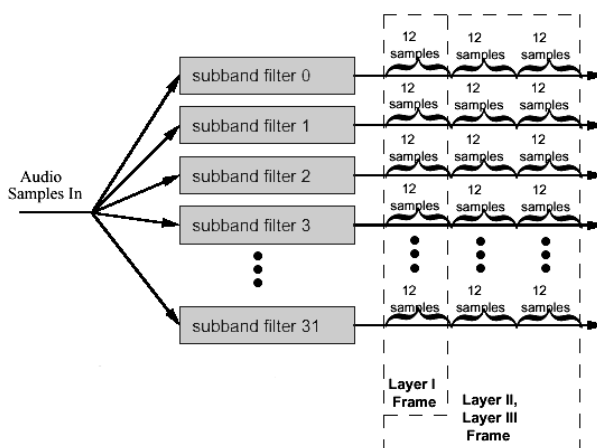


Bild 2.15: Filterbank der MPEG-Encoder

muß (*Bit Allocation*) und andererseits, welches Teilband gar nicht übertragen werden muß, entsprechend zu Bild 2.12 (unten). Diese Zusatzinformationen (*side information*) muß übertragen und dem Decoder mitgeteilt werden, damit dieser das Audiosignal dann wieder rekonstruieren kann.

Um die Verdeckungskurven zu bestimmen, führt das psychoakustische Modell (parallel zu der Analyse mit den Subband-Filtern) eine FFT mit 1024 Punkten durch, wodurch sich eine frequenzmäßige Auflösung von ≤ 20 Hz ergibt. Für die verdeckten (maskierten) Anteile werden die Samples der betreffenden Subband-Filter zu 0 gesetzt (entsprechend zu Bild 2.12 unten) und damit nach einer Huffman-Codierung nicht übertragen.

3 Kaskadierung von Strecken mit MPEG-Codierung

Bei den Zubringer-Strecken für Audio-Signale im Rundfunk haben die Schnittstellen am Eingang und Ausgang jeder Teil-Strecke das bei der CD übliche Format (16 Bit linear quantisiert, 44,1 KHz), sind also zur CD kompatibel. Auf dem Streckenabschnitt wird jedoch eine Datenreduktion bzw. Kompression nach MPEG2 durchgeführt, um Übertragungsbandbreite zu sparen. Da jede Teil-Strecke zur CD kompatibel ist, können auch mehrere Teilstrecken auf einfache Weise kaskadiert werden.

Vom Rundfunkstudio bis zum Sender können so leicht 4 bis 5 solcher Teilstrecken zusammenkommen. Mittlerweile sind auch die Schallarchive des Rundfunks weitestgehend digitalisiert und MPEG2 codiert, was formal als weitere Teilstrecke betrachtet werden kann.

Die beim Rundfunk üblichen Datenraten sind 384 kBit/s, 256 kBit/s und 192 kBit/s, wobei das Archiv 384 kBit/s verwendet. Das entspricht Kompressions-Raten von 1:4 (384 kBit/s) bzw. 1:6 (256 kBit/s) bzw. 1.8 (192 kBit/s). Alle diese Kompressionsraten sind so, daß das normale Ohr einen Unterschied zum nicht komprimierten Signal nicht bemerkt.

In einer Untersuchung, deren Ergebnisse 1995 vorgestellt wurden [2], werden die Auswirkung von kaskadierten MPEG2 codierten Teilstrecken auf die Hörbarkeit von Artefakten beschrieben.

Untersucht wurden zwei Kaskadierketten mit den (anonymisierten) Codecs A, B, C, D, E, F:

1. 384 kBit/s — 256 kBit/s — 384 kBit/s — 256 kBit/s
2. 384 kBit/s — 256 kBit/s — 384 kBit/s — 192 kBit/s (js)^{†1}

Es wurden dabei alle Teil-Strecken mit jeweils dem selben Codec ausgeführt. Als Testsignale fanden „dire straits“, „harpsicord“, „stravinsky“ und „speech“ Verwendung.

Zusammenfassend wird festgestellt:[2]

„Die Ergebnisse zeigen, daß bei verschiedenen Testsignalen zum Teil große Unterschiede zwischen den einzelnen ISO/MPEG - Codecs bestehen, die daher stammen, daß die Encoderimplementierung nicht genormt und damit bei den einzelnen Realisationen verschieden ist. Man kann deutlich erkennen, daß einige

^{†1}Joint Stereo (js) bildet das Summen- und das Differenz-Signal, wie das bei der Stereo-Übertragung im UKW FM Rundfunk üblich ist. Im Differenz-Signal steckt weniger Energie als im Summen-Signal. Das Summen-Signal wird daher genauer codiert als das Differenz-Signal, wodurch insgesamt Datenrate gespart werden kann.

Hersteller sich stärker mit der Audioqualität befaßt haben als andere. Für alle gilt aber, daß in Codierverfahren wohl noch einige Ressourcen stecken, da die Prototypen zum großen Teil eine bessere Qualität besitzen als die am Markt erhältlichen Codecs.

Insgesamt aber erreichen die meisten Codecs bei den ausgewählten Ketten nur bei einem relativ unkritischen Signal wie „dire straits“ Rundfunkqualität und auch noch einige bei „speech“. Bei Signalen wie „stravinski“ oder dem sehr kritischen „harpischord“ muß jedoch noch einiges an der Implementierung der Codecs verbessert werden, um Rundfunkqualität zu gewährleisten. [Stand: 1995]

Außerdem fällt auf, daß sich die Codecs auf bestimmtes Audiomaterial, wie z. B. Sprache, optimieren lassen, und dann bei anderem Material, wie z. B. Popmusik, unter Umständen nicht die gewünschte Qualität erreichen. Generell sollte immer mit der höchsten Datenrate gefahren werden, die möglich ist, um jegliche Qualitätsverschlechterung durch Kaskadierung zu vermeiden.^{‡2}

Die Kaskadierung von Strecken mit Geräten von unterschiedlichen Produzenten, wie es in der Praxis üblich ist, birgt ebenfalls das Problem, daß sich die psychoakustischen Modelle unterscheiden, da diese ja firmengeheim sind. Demzufolge können bei einer Kette mehr spektrale Komponenten unterdrückt oder verändert werden, als es bei einer einzelnen Teilstrecke der Fall ist. Das führt dann auf eine Erhöhung der effektiven Kompressionsrate entsprechend zu einer Reduzierung der effektiven Datenrate, die dann ggf. nicht mehr im Einklang mit dem definierten psychoakustischen Modell steht. Und Personen mit den „goldenen Ohren“ können dann Unterschiede wahrnehmen.

4 Audioprocessing am UKW-Sender

Während die MPEG Codierung durch die Berücksichtigung des psychoakustischen Modells bestrebt ist, das originale Audio-Signal bezogen auf den Höreindruck möglichst wenig zu verfälschen, wird direkt am Sender das Audiosignal einer Transformation unterworfen mit dem Ziel, es möglichst „laut“ zu machen.^{§1} Als Gerät hat sich hier der „Optimod“ von Orban etabliert. Möglichst laut bedeutet:

1. Verminderung der Dynamik auf Werte ≤ 3 dB.
2. Phasendrehung sämtlicher Obertöne so, daß der Crestfaktor minimal wird.

Durch die Reduzierung der Dynamik klingt alles praktisch gleich laut, Bild 4.1.

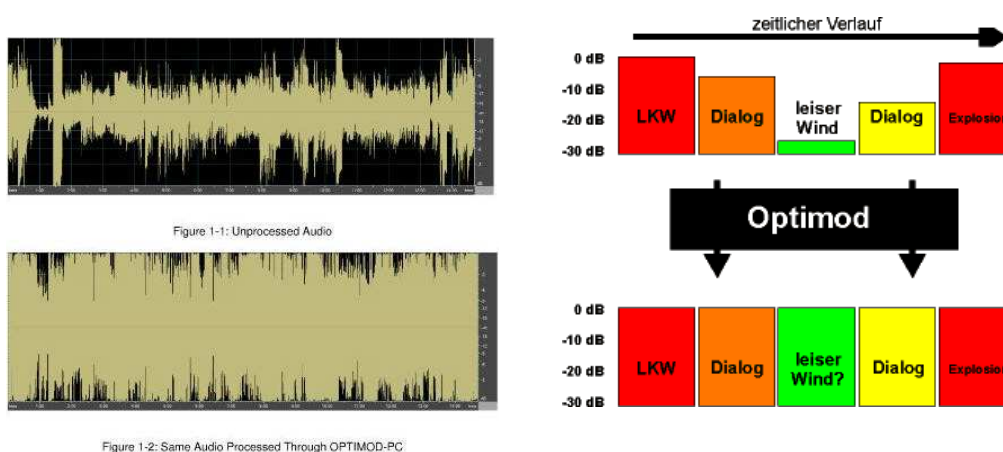


Bild 4.1: Reduktion der Dynamik eines Audiosignals (links), schematisch (rechts)

Die Reduzierung der Dynamik macht durchaus Sinn, wenn Radio in einer geräuschbehafteten Umgebung gehört wird, wie z.B. im Auto. Leise Passagen gingen da einfach „unter“, wenn ihr Pegel nicht angehoben wäre.

Der zweite Punkt betrifft eine Eigenschaft des FM-Rundfunks. Hier gibt es eine Begrenzung des zulässigen Hubes auf ± 75 KHz. Nun ist die Größe des (Spitzen-) Hubes direkt proportional zu der maximalen Amplitude des modulierenden Signals. Also erscheint es sinnvoll, das NF-Signal so zu bearbeiten,

^{‡2}Nach Auskunft von einem der Autoren gelten die Aussagen i.W. heute auch noch.

^{§1}Es gibt einzelne Ausnahmen von dieser Praktik, so z.B. die Klassik-Programme der ARD und DRadio (DLF und DRadioKultur).

daß zwar alle Frequenzkomponenten erhalten bleiben, aber die maximale Amplitude des NF-Signals minimiert wird. Dann kann das NF-Signal insgesamt „lauter aufgedreht“ werden, ohne den Spitzen-Hub zu überschreiten.^{§2§3} Da das Ohr i.W. unempfindlich bezüglich der Phasendrehungen der Oberschwingungen ist, fällt diese Manipulation nicht auf.

Über die Realisierung des Optimod gibt es keine detaillierte Information. [15] Aufgrund des Standes der Technik ist jedoch sehr wahrscheinlich, daß das Signal in kurze Zeitabschnitte zerlegt wird und bei jedem wieder eine FFT durchgeführt wird um die spektralen Komponenten zu bestimmen. Dann lassen sich in einfacher Weise die Pegel anheben und die Phasen der Obertöne drehen. Anschließend erhält man über eine IFFT wieder die Abschnitte des Zeitsignals, die dann zusammengefügt werden.

Der Optimod wurde schon zu Zeiten eingesetzt, als die Zubringer-Leitungen zu den Sendern noch weitestgehend analog waren. Da analoge Leitungen i.a. keine lineare Phase haben, war es nicht möglich, bereits am Studio-Ausgang einen Optimod einzusetzen, da dann beim Sender die Phasen der Oberschwingungen so weit gedreht waren, daß der Crestfaktor des Signals nicht mehr minimal war. Folglich hat jeder UKW-Sender, bei dem Optimod eingesetzt wird, ein solches Gerät direkt am Modulationseingang.^{§4}

Bei digitaler Zuführung des Audiosignals für UKW-Sender über Satellit könnte heute auch ein (einzelner) Optimod am Anfang der Satelliten-Strecke eingesetzt werden. Im allgemeinen ist das bei Satellitenprogrammen (DVB-S) heute noch nicht der Fall so daß beim digitalen Empfang eines Satelliten das Audio-Signal weniger „verfremdet“ ist als bei manch einem UKW-Sender.

Trotz Allem: Ausgewogener Klang im Röhrenradio Wenn der Empfang eines UKW-Programms, dessen Audio-Signal heute alles andere als „naturgetreu“ ist, mit einem Röhren-Radio aus den 50er/60er Jahren trotzdem „angenehm“ klingt, liegt das an dem als ausgewogen empfundenen Klang dieser Geräte. Die waren zwar meist nicht „HiFi“. Aber das eine sind die Meßwerte und das andere sind (subjektive) Klang-Empfindungen.

^{§2}Im Spektrum der FM ist die Bandbreite nur näherungsweise proportional zum Hub ΔF . Gemäß Carson ist $B_{FM} > \approx 2 \cdot (\Delta F + 2f_{\max})$, wenn f_{\max} die höchste Modulations-Frequenz ist. Die zulässige Bandbreite bei FM wird daher bei Optimod prozessierten Signalen überschritten, wenn nur auf den Spitzenhub geachtet wird.

^{§3}Es gibt auch Optimod für AM-Sender. Hier gilt es dann, Übermodulation zu verhindern und trotzdem nahe an 100% Modulationsgrad heran zu kommen.

^{§4}Speziell bei privaten Sendern wird der Optimod auch dazu verwendet, einen typischen (unverwechselbaren ?) Klang zu erzeugen.

Literatur

- [1] Bahr, H.: *Philips Lehrbriefe, Elektrotechnik und Elektronik*, Bd. 1, Hüthig, 1982
- [2] Felderhoff, U.; Ritscher, S.: *Kaskadierung verschiedener Audiodatenreduktionssysteme*, ITG Fachbericht „Hörrundfunk“ 133, 1995, pp. 177 ff.
- [3] Oppenheim, A.V.; Schaffer, R.W.: *Discrete-Time Signal Processing*, Prentice Hall, 1989
- [4] Bergmann, K. (Hrsg.): *Lehrbuch der Fernmeldetechnik*, 5.A., Schiele & Schön, 1986
- [5] Haykin, S.: *Communication Systems*, 3rd ed., Wiley, 1994
- [6] Trendelenburg, F.: *Einführung in die Akustik*, 3.A., Springer, 1961
- [7] Feldkeller, R.; Zwicker, E.: *Das Ohr als Nachrichtenempfänger*, Hirzel, Stuttgart, 1956
- [8] Zwicker, E.; Feldkeller, R.: *Das Ohr als Nachrichtenempfänger*, 2.A., Hirzel, Stuttgart, 1967
- [9] Dambacher, P.: *Digitale Technik für Hörfunk und Fernsehen*, Decker, 1995
- [10] Rudolph, D.: *Anwendungen der Fourier-Transformation*, Kap. 5. www.diru-beze.de
- [11] o.A. (Hewlett-Packard): *MPEG-2: The Basis of how it works*, 6C06Mpegtutorial1.pdf
- [12] Weißflog, J.: *Verbesserungen am ISO – MPEG1 – Layer2 – Encoder*, ITG Fachbericht „Hörrundfunk“ 133, 1995, pp. 171 – 175
- [13] Rudolph, D.: *Frequenz, Spektrum, Resonanz*, www.diru-beze.de
- [14] Edler, B.: *Prädiktive Teilbandcodierung mit Vektorquantisierung für Audiosignale hoher Tonqualität*, ITG Fachbericht „Hörrundfunk“ 106, 1988, pp. 223 – 228
- [15] Orban, R.; Ogonowski, G.: *Maintaining Audio Quality in the Broadcast Facility*, Orban, 2003