

Seminar Neue Medien

an der

Universität Trier

Leitung: Prof. Bernd Walter, Dipl. Inf. Alexander Weber

Prinzipien der Psychoakustik

Schriftliche Ausarbeitung zum Seminarvortrag am 11.06.2002

Autor: Jörg Follmann, Matr. Nr.: 601561

INHALTSVERZEICHNIS ZUM**SEMINAR: NEUE MEDIEN PRINZIPIEN DER PSYCHOAKUSTIK****UNIVERSITÄT TRIER****LEITUNG: PROF. BERND WALTER, DIPL. INF. ALEXANDER WEBER**

1.	MOTIVATION	3
1.1	Einführendes Beispiel	3
1.2	Aufbau der Arbeit	3
1.3	Typisches Vorgehen bei der Audiocodierung	4
1.4	Grundlagen der Digitalisierung von Musik	4
1.4.1	Sample und Quantisierung	4
1.4.2	Pulse Code Modulation	5
2.	EIGENSCHAFTEN DES MENSCHLICHEN GEHÖRS	5
2.1	Hörschwellen	5
2.2	Frequenzbänder	6
2.3	Wahrnehmungseigenschaften	6
3.	METHODEN IN DER PSYCHOAKUSTIK	6
3.1	Maskierungseigenschaften	6
3.2	Simultane Maskierung	7
3.2.1	Noise – Masking – Tone	7
3.2.2	Tone – Masking – Noise	8
3.2.3	Asymetrie in der Simultanen Maskierung	9
3.2.4	Noise – Masking – Noise	9
3.3	Zeitliche Maskierung	9
3.3.1	Vorwärtsmaskierung	10
3.3.2	Rückwärtsmaskierung	10
3.4	Reduzierung der Stereo - Informationen	10
4.	PRAKTISCHE UMSETZUNG IN ENCODERN	11
4.1	Frequenzbänder	11
4.2	Simultane Maskierung	11
5.	BEISPIELE UND DATEN	12
5.1	Beispiel MPEG-1	12
5.2	Daten für ausgewählte Audioverfahren	13
5.3	Effizienz der vorgestellten Verfahren	13
	ABBILDUNGS - UND QUELLENVERZEICHNIS	15

1. Motivation

1.1 Einführendes Beispiel

Heute werden immer mehr Audio – Daten in digitaler Form weitergeleitet. Um den Datenverkehr bei einer Internet – Übertragung oder den Speicherplatz zur Speicherung von Audio – Formaten zu minimieren, ist es notwendig, analog zur Bildbearbeitung und den dort eingesetzten Kompressionsverfahren, effiziente Methoden zur verlustbehafteten Audio – Codierung zu entwickeln.

Da es sich hier um Kompressionsverfahren, nicht um Komprimierungsverfahren, die für eine ausreichende Senkung der Übertragungsraten nicht geeignet sind, handelt, ist ein Informationsverlust die logische Konsequenz. Dieser soll aber so gering wie möglich gehalten werden, d.h. hier, so wenig wie möglich störend auf den Hörer wirken und vorrangig unnütze Daten aus dem Audio – Strom extrahieren. Das Ziel dieses Reduktionsvorganges ist durch Verwendung psychoakustischer Methoden, einen möglichst guten Kompromiss zwischen hoher Klangqualität und minimaler Dateigröße, bzw. Übertragungsrate zu erzielen.

Moderne Verfahren wie z. B. der MPEG – Standart erreichen Komprimierungen von mit bis zu 12 – facher Verkleinerung, bei keinem bis wenig Qualitätsverlusten.

Beispiel(1):

(Die verwendeten Begriffe werden in Kapitel 1.4 definiert, die zur Berechnung verwendeten Angaben beziehen sich auf die in Kapitel 5.2)

Bei der Speicherung einer Minute Musik in CD – Qualität auf einer Festplatte, wird die CD mit einer Frequenz von 44,1 kHz bei einer Quantisierung von 16 Bit pro Sample abgetastet. Daraus ergibt sich folgende Rechnung:

$$44,1\text{kHz} * 16 \text{ Bit} * 2 = 1411,2\text{kBit/sec. als Datenrate.}$$

Der Faktor 2 beruht darauf, dass 2 Kanäle abgetastet werden (Stereo). Dies entspricht für die gefragte Minute ca. 10MB Speicherplatz.

1.2 Aufbau der Arbeit

Alle erwähnten Methoden und Verfahren der Psychoakustik beruhen auf empirischen Untersuchungen an Versuchspersonen. Die ermittelten Ergebnisse wurden für die Definition verschiedener psychoakustischer Modelle verwendet. Psychoakustik ist keine mathematisch oder physikalisch beweisbare Wissenschaft, wenngleich hier die wichtigsten physikalischen Eigenschaften des menschlichen Hörvermögens für die Definition der Modelle ausgenutzt werden. Diese wichtigen Hör – Eigenschaften werden in Kapitel 2 näher erläutert und beschrieben.

Die zur psychoakustischen Codierung verwendeten Methoden werden in Kapitel 3 detailliert vorgestellt. Die Verwendung einiger dort vorgestellter Methoden in Audio – Encodern wird in Kapitel 4 detailliert erläutert. In Kapitel 5 folgt als Beispiel die genauere Betrachtung des psychoakustischen Vorgehens beim MPEG-1 – Standart, sowie einige ermittelte Ergebnisse der Audio – Codierung, wie Datenraten oder Kompressionsfaktoren

bei unterschiedlichen Verfahren. Zunächst soll jedoch eine kurze Vorstellung von typischen Verfahren im Audio - Bereich folgen:

1.3 Typisches Vorgehen bei der Audiocodierung

Um die Rolle der Psychoakustik bei der Audio – Codierung besser einordnen zu können, wird an der folgenden schematischen Darstellung eines generischen Audio – Encoders die Arbeitsweise der meisten Codierungsalgorithmen kurz dargestellt:

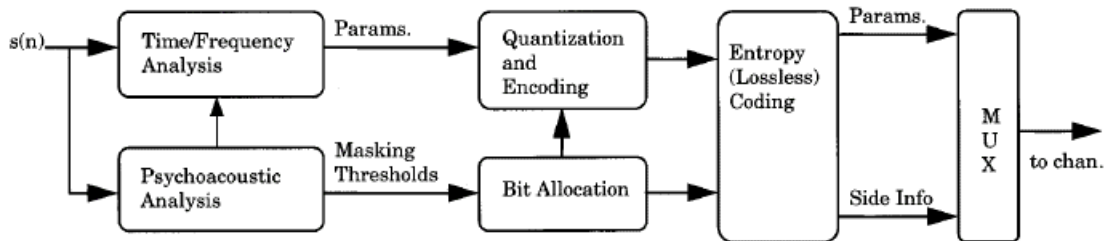


Abb. 1: Generisches Modell eines Audio – Encoders

Zunächst wird das Audio – Signal in konstante zeitliche Einheiten von 2 – 50msec, abhängig des jeweils ausgewählten Codier - Verfahrens, in „Frames“ zerhackt. Parallel zur zeitlichen Analyse und der Frequenzanalyse wird ein psychoakustisches Modell über den Eingabestrom gelegt. Die hier erzielten Ergebnisse, wie Verdeckungs – und Hörschwellen werden anhand aufwendiger Berechnungen (z. B. wird folgendes Verfahren zur weiteren Komprimierung angewendet: Benötigt ein Block weniger als die Anzahl an Bits, wie durch die Quantisierung vorgesehen sind, so werden die übrigen Bits an das sogenannte "Bit – Reservoir"¹ übergeben. Sind nun in einem nächsten Quantisierungsschritt mehr Bits als vorgesehen notwendig, können diese durch das vorhandene Reservoir ausgeglichen werden.) für einen weiteren, diesmal verlustfreien Codierungsschritt vorbereitet. In vielen Modellen, wie z. B. auch bei MP3, wird dazu das Verfahren des Huffman – Coding eingesetzt. In diesem Vortrag werden diese Methoden zur verlustfreien Codierung keine Rolle spielen.

1.4 Grundlagen der Digitalisierung von Musik

1.4.1 Sample und Quantisierung

Beim Digitalisieren eines analogen Audio – Signals werden die physischen Schwingungen auf ganze Zahlen abgebildet, die **Samples**, die meist in 16 – Bit – Wörtern dargestellt werden. Dieser Vorgang wird **Sampling** genannt. Die Länge der Wörter heißt im Deutschen **Abtastwert (Quantisierung)**. Notwendigerweise entstehen beim Runden dieser Zahlen Fehler, sogenannte **Quantisierungsfehler**, die maximal ein Bit betragen und später als **digitales Rauschen** oder **Quantisierungsrauschen** bezeichnet werden.

Die **Sampling - Frequenz** (engl.: **sampling rate**, kurz: **sr**) wird in Hertz angegeben. Sie gibt an wie viele Samples pro Sekunde abgetastet werden.

Die im Beispiel(1) in Kapitel 1 berechnete Datenrate ergibt sich also aus:

¹ Aus Quelle[4] genaue WWW – Seite: <http://www.tecchannel.de/multimedia/57/4.html>

Sampling – Frequenz * Quantisierung/sec * #Kanäle

1.4.2 Pulse Code Modulation

Zur Digitalisierung im konventionellen Bereich, also bei Audio – CD und DAT wird als grundlegendes Verfahren die **Pulse Code Modulation (PCM)** mit einer Abtastfrequenz 44,1 kHz und einer Quantisierung von 16 Bit pro Sample verwendet. Man unterscheidet zwischen **dynamischen** und **linearen** Verfahren. Letzteres verwendet einheitliche Quantisierungsschritte mit dem Nachteil, dass abhängig vom aktuellen Schalldruckpegel, stark abweichendes Rauschen in Kauf genommen werden muss. Durch die starken Abweichungen fallen hohe Datenmengen an.

Bei den dynamischen Verfahren werden die Quantisierungsschritte so angepasst, dass keine großen Schwankungen beim Rauschen auftreten, was auch die zugehörigen Datenmengen weniger stark beeinflusst.

2. Eigenschaften des menschlichen Gehörs

In diesem Kapitel werden nur diejenigen Eigenschaften diskutiert, die auch für das Verständnis der weiteren Kapitel wichtig sind.

2.1 Hörschwellen

Ein gesundes (junges) menschliches Ohr hat einen Hörbereich von 16 Hz – 20 kHz². Hieraus folgt sofort, dass Signale außerhalb des Hörbereichs bei der Codierung nicht relevant sind.

Nicht nur die **Frequenz** (Anzahl der Schwingungen pro Zeiteinheit), sondern auch der **Schalldruckpegel** (engl.: Sound Pressure Level, **SPL**) eines Tons sind entscheidend für seine **Hörbarkeit**.

Die **Ruhehörschwelle** gibt an ab welchem Schallpegel das Ohr einen reinen Sinuston hören kann. Sie ergibt sich als Kurve in Abhängigkeit der Frequenz. Aus dieser Funktion ist ersichtlich, dass der empfindlichste Hörbereich des Menschen zwischen 2500 und 5000 Hz liegt. Hier ist nur ein geringerer Schalldruckpegel zur Erregung des Hörorgans als außerhalb des empfindlichsten Hörbereichs notwendig.

Der Lautstärkepegel der Umgebung erhöht die Hörschwelle und **maskiert** leisere Töne. Man nennt den notwendigen minimalen Schalldruckpegel eines Tons, der notwendig ist, um ihn neben dem **Maskierer** (Nebengeräusch) wahrnehmen zu können, **Mithörschwelle**.

Wichtiger als die oben angesprochenen Hörschwellen ist die spektrale Analyse der wahrgenommenen Signale, die wie im folgenden Kapitel beschrieben modelliert werden kann.

² Aus Quelle[2]

2.2 Frequenzbänder

Um die Höreigenschaft des menschlichen Ohrs möglichst genau und einfach zu beschreiben, wird folgendes Modell benutzt:

Eintreffende komplexe Schallspektren zerlegt das menschliche Ohr durch eine Art Fast – Fourier – Transformation in ca. 24 (Quelle [1] spricht von 25, mehrere andere Quellen, wie z. B. [2] von 24 kritischen Bandbreiten) aneinandergereihte, teilweise überlappende **Frequenzgruppen** oder Bandbreiten. Die dadurch entstehende Skala wird nach dem Physiker Heinrich Georg Barkhausen (1881 – 1956) als **Barkskala** bezeichnet. Zu Bemerkem ist, dass diese Skala nicht linear ist, sondern dass die Bandbreiten im höheren Frequenzbereich eine größere Ausdehnung haben. So hat ein Frequenzband mit einer Frequenzmitte von 50 Hz eine absolute Breite von 100 Hz, ein Frequenzband mit der Mitte von 1000 Hz schon eine Bandbreite von 920 Hz – 1080 Hz, also eine absolute Bandbreite von 160 Hz. Am Modell ebenfalls leicht einzusehen ist, dass durch Reizung eines Bandes ein weitere, benachbarte Bänder ebenfalls in Schwingungen versetzt werden können. Dieses Phänomen wird Kapitel 3.2 als „*Spread of Masking*“³ besprochen.

Eintreffende Schallwellen versetzen bestimmte Bereiche dieser Bänder in Schwingungen. Hierfür braucht das menschliche Ohr eine gewisse Ein – bzw. Ausschwingphase, besonders bei sehr lauten oder leisen Tönen. Ausgenutzt wird dieses Phänomen bei der *Zeitlichen Maskierung* (s. Kapitel 3.3).

Die empfundene Lautstärke eines Geräuschs, dass mit konstantem SPL anliegt, erhöht sich erst nachdem das Geräusch die kritische Bandbreite überschritten hat. Werden so zwei oder mehrere benachbarte Bänder in Schwingung versetzt, steigt die empfundene Lautstärke an.

2.3 Wahrnehmungseigenschaften

Eine weitere wichtige Höreigenschaft bei der Codierung ist die Unterscheidung in periodische (Sinustöne) und aperiodische (Rauschen, Trommeln) Signale (s. Kapitel 3.1, 4.2).

Als letzte Höreigenschaft soll nun noch kurz das Prinzip für die Aufnahme von Stereo – Signalen erläutert werden.

Das menschliche Gehirn registriert den unterschiedlichen Zeitpunkt der Aufnahme eines Schallsignals durch das linke bzw. rechte Ohr und berechnet hieraus, wie aus den unterschiedlichen Schalldruckpegeln links, bzw. rechts, die Position der Schallquelle. Diese Informationen werden bei der Aufnahme von Musik in Stereo – Qualität entsprechend berücksichtigt und kosten entsprechende Kapazitäten.

3. Methoden in der Psychoakustik

3.1 Maskierungseigenschaften

Die größten Reduzierungsmöglichkeiten zeigen Verdeckungen (Maskierungen) auf, die anhand psychoakustischer Modelle ermittelt werden. So können Informationen, deren

³ Aus Quelle[1] Seite 459

Wiedergabe durch andere dominantere und damit „überdeckende“ Signale nicht wahrnehmbar ist, schon bei der Codierung aus dem Audio – Strom extrahiert werden. Sie sind für die Qualität der Sound – Wiedergabe nicht relevant. Die Methoden zur Klassifizierung in relevante und nicht relevante Informationen werden in den folgenden Kapiteln des 3. Abschnitts beschrieben.

Generell gilt, ein kurzes, tonales Signal ist stärker verdeckend als geräuchhafte, breitbandige atonale Signale. Ein Maskierer höherer Frequenz kann nur einen Schall mit geringem Frequenzabstand verdecken. Andererseits kann ein tiefer Maskierer einen höherfrequenten Schall nur mit einem vergleichsweise hohem Schallpegel verdecken.

Bei der Audio – Codierung werden in einem ersten Analyse – Schritt Geräusche und reine Töne unterschieden. Mathematisch sind Geräusche von tonalen Signalen dadurch unterscheidbar, dass die Verlaufskurve eines Tones Sinus – artig ist, die eines Geräusches nicht vorhersehbar ist.

Moderne Audio – Encoder unterscheiden zwischen Simultaner und Nicht – Simultaner (zeitlicher) Maskierung. Simultane (3.2) und Zeitliche (3.3) Maskierung reichen zu einer Reduzierung um den Faktor 5.

3.2 Simultane Maskierung

Simultane Maskierung oder Verdeckung setzt die Anwesenheit eines dominierenden Geräuschs oder Tones voraus, das bzw. der eine ausreichende Erregung eines kritischen Bandes erreicht, so dass jedes schwächere Signal auf dem gleichen Band oder seiner Umgebung nicht wahrgenommen wird. Die **Signal – to – Mask – Ratio**, SMR, ist definiert als die Differenz der SPL des Maskierers und der SPL des maskierten Signals und wird ebenfalls in Dezibel (dB) angegeben.

Aus der Unterscheidung von Tönen und Geräuschen ergeben sich im Wesentlichen drei voneinander abgrenzbare Phänomene bei der simultanen Maskierung: Ein dominantes Geräusch verdeckt einen reinen **Sinus – Ton**, im Folgenden oft nur Ton genannt, (**Noise – Masking – Tone**, NMT), ein dominanter Ton überlagert ein Geräusch (**Tone – Masking – Noise**, TMN) und ein Geräusch verdeckt ein weiteres (**Noise – Masking – Noise**, NMN). Diese drei Phänomene werden nun vereinfacht diskutiert, denn im Vorhinein sei gesagt, dass die Maskierungseffekte nicht wie im Folgenden dargestellt, auf ein kritisches Band (1Bark) beschränkt sind, sondern wie schon in Kapitel 2.2 angesprochen, dass eingehende Signale auch angrenzende Bandbreiten reizen können. Auch ein (wie in den folgenden Kapiteln angenommen) über der kritischen Bandbreite zentriertes Signal hat Auswirkungen auf andere Bänder. Dieser Effekt wird im Englischen als „**Spread of Masking**“ bezeichnet.

3.2.1 Noise – Masking – Tone

Bei NMT maskiert ein Geräusch einen Ton im gleichen kritischen Frequenzband dadurch, dass die Intensität des Reizes ausgehend vom Ton wesentlich geringer ist als die des maskierenden Geräuschs. Die SMR ist minimal, wenn die Frequenzmitte des Geräuschs und die Frequenz des maskierten Tons nahe beieinander liegen, bzw. gleich sind, wie im Beispiel (2). In vielen Studien hat die minimal erreichbare SMR einen Betrag von 4 dB. Wird die Frequenz des Tons erhöht oder erniedrigt steigt die SMR zwischen Ton und Geräusch.

Beispiel(2):

Ein reiner Ton mit einer Frequenz von 410 Hz und einem SPL von 76 dB wird von einem band – nahen Geräusch (Frequenzbreite: 1Bark), zentriert um 410 Hz mit einer Bandbreite von 90 Hz und 80 dB SPL, verdeckt.

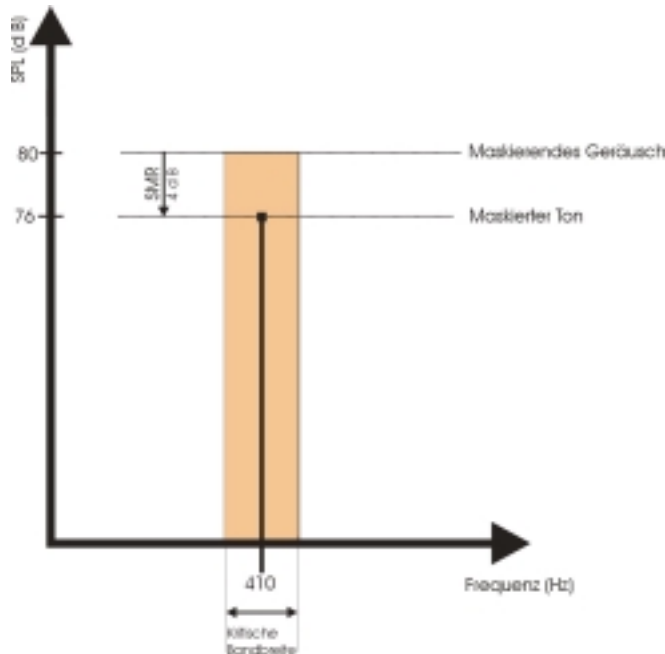


Abb. 2: Beispiel (2): NMT

3.2.2 Tone – Masking – Noise

Ein einzelner Ton mit einer Frequenz f maskiert ein Geräusch, dessen Frequenzmitte eines kritischen Bandes f ist, vorausgesetzt dass sein SPL entsprechend über dem des Geräusches liegt. Aus Untersuchungen zum TMN, die übrigens weniger zahlreich als die zum oben erklärten NMT sind, folgt, dass die minimale SMR zur Verdeckung eines Geräusch durch einen Ton zwischen 21 und 28 dB liegt.

Auch hier steigt die SMR an, wenn sich der verdeckte Ton von der Frequenzmitte des Geräusch entfernt.

In einigen Fällen wird das TMN – Phänomen auch ausgenutzt um das im Kapitel 1.4.1 besprochene Quantisierungsrauschen für das menschliche Ohr unhörbar zu machen, bzw. das Quantisierungsrauschen wird erst gar nicht unterdrückt oder verhindert, da es wegen der verdeckenden Eigenschaft des Tones vom Gehör nicht wahrgenommen wird.

Beispiel(3):

Die Grafik zeigt ein Geräusch zentriert über der Frequenz eines Tons von 1000 Hz. Die SPL des Tons beträgt 80 dB, die des maskierten Geräuschs 56 dB. Daraus ergibt sich die minimale SMR von 24 dB in diesem Fall.

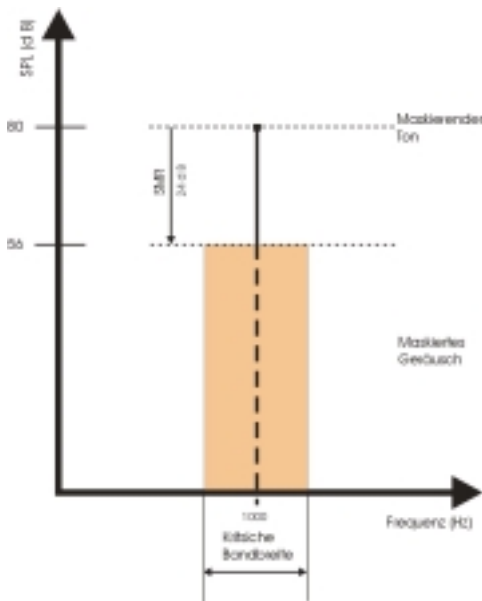


Abb. 3: Beispiel (3): TMN

3.2.3 Asymmetrie in der Simultanen Maskierung

Vergleicht man die Beispiele (2) und (3) zeigt sich eine gewisse Asymmetrie zwischen NMT und TMN. Leicht zu sehen ist, dass die minimale SMR im Fall von TMN höher liegt als im vorher betrachteten Fall (NMT), d. h. ein einzelner Ton als Maskierer benötigt eine weitaus höhere SPL als das zu maskierende Geräusch, um es zu verdecken als umgekehrt. Während die Differenz

$$SMR_3 - SMR_2 \quad \text{wobei } SMR_3 \text{ die SMR aus Beispiel(3) und } SMR_2 \text{ die SMR aus Beispiel(2) ist.}$$

ungefähr 20 dB beträgt ist in beiden Fällen die SPL des Maskierers gleich.

3.2.4 Noise – Masking – Noise

Dieses Szenario wurde nur in sehr wenigen Untersuchungen betrachtet. Ein wichtiger Aspekt beim Filtern eines der beiden Geräusche ist, darauf zu achten, dass beide Geräusche sich unter Umständen mischen können. Die Zuordnung zum zugehörigen Geräusch fällt schwer, denn hier ist eine Trennung in periodischen und aperiodischen Anteil nicht möglich. Deshalb bedeutet diesen Effekt auszunutzen eine genaue Untersuchung des Eingangssignals.

3.3 Zeitliche Maskierung

Wegen der im Kapitel 2.2 angesprochenen Ein – und Ausschwingphase des Ohrs treten neben der simultanen auch zeitliche oder nicht – simultane Verdeckungen auf. Man unterscheidet **Vorwärts – und Rückwärtsmaskierung**.

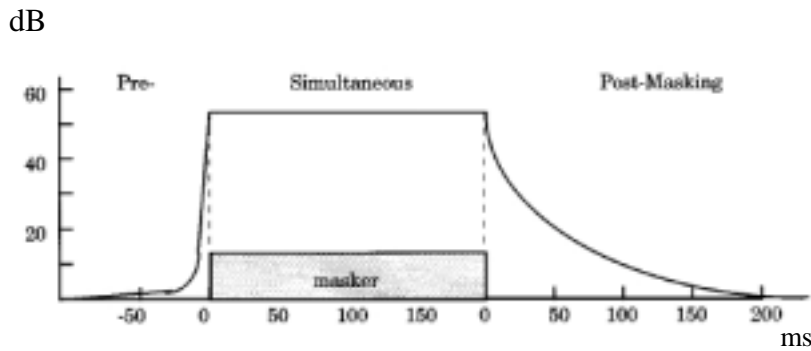


Abb. 4: Vor – und Rückwärtsmaskierung

Wie in der Abbildung zu erkennen ist die für die Reduzierung wichtigere die Vorwärtsmaskierung oder Postmaskierung, da die Beeinflussung des Gehörs noch 50 – 200 msec nach Beendigung des eigentlichen Signals reicht und deshalb hierbei eine erheblich höhere Kompressionsrate erzielt werden kann als bei der Rückwärtsmaskierung oder Premasking von höchstens 1 – 2 msec Dauer.

Außer der Zeitdauer des anliegenden Signals sind Frequenz, Schalldruckpegel und die Art des Signals (periodisch oder aperiodisch) für die maximale Maskierungsdauer ausschlaggebend.

3.3.1 Vorwärtsmaskierung

Nach lauten Tönen erlangt das Gehör erst allmählich wieder die ursprüngliche Hörempfindlichkeit. Die Mithörschwelle sinkt nach der Wahrnehmung eines lauten Signals kontinuierlich, jedoch nur langsam ab. Die Zeitdauer des Absinkens hängt von den oben beschriebenen Faktoren des ausschlaggebenden Schallereignisses ab. Beispielsweise würde ein kurzer Trommelschlag eine schneller abfallende Nachmaskierungsschwelle hervorrufen, als ein lauter dröhnender Bass - Ton.

Für die Mithörschwelle im nächsten Frame folgt nun aus psychoakustischer Sicht folgendes: Wurde im aktuellen Frame eine extrem hohe Mithörschwelle berechnet und wäre im folgenden Frame für sich genommen eine niedrigere Mithörschwelle gültig, so wird diese hier von der Mithörschwelle des vorherigen Frames nach oben gesetzt.

3.3.2 Rückwärtsmaskierung

Signale werden vom Gehör in Abhängigkeit von Schalldruckpegel und Frequenz mit Verzögerung wahrgenommen. Deshalb ist es möglich, dass ein lautes, kurzes Signal, beispielsweise ein Trommelschlag, ein gleichzeitig anliegendes leises Geräusch, bzw. die zugehörigen jeweiligen Frequenzen, „überholen“ kann und so zuerst vom Ohr wahrgenommen werden. Zusammen mit den in den vorigen Kapiteln vorgestellten Überdeckungseigenschaften, ist das leise Geräusch also insgesamt überhaupt nicht hörbar. Auch diese Eigenschaft kann beim Überdecken des dauernd anliegenden Quantisierungsrauschens ausgenutzt werden.

3.4 Reduzierung der Stereo - Informationen

Stereo – Informationen sind für die Hörqualität von Musik sehr wichtig. Stereo – Informationen zu filtern, kann einerseits einen hohen Grad an Kompression erfüllen, andererseits, Musik, im Gegensatz zur Sprache, fast zerstören. Deshalb eignen sich Verfahren der Reduzierung der Stereo – Informationen überwiegend in Bereichen, in

denen schnelle Übertragung von Sprache gefordert ist. Im günstigsten Fall erreicht man durch das nachfolgend vorgestellte Verfahren des „Joint – Stereo“⁴, eine Kompression um den Faktor fünf.

Das Verfahren nutzt zwei Phänomene:

- a) Tiefe Töne sind vom menschlichen Gehör nicht zu lokalisieren. Es reicht also für tiefe Frequenzen aus, sie ausschließlich über einen Kanal zu übertragen, und sie auf dem parallelen Kanal zu filtern. Diese Technik ist vom verbreiteten „Subwoofer – System“ bekannt.
- b) Bei der Übertragung von Stereo – Informationen werden neben zwei Seitenkanälen, auch Informationen über einen Mittenkanal weitergegeben. Dieser enthält deutlich weniger Informationen (wegen geringerem Schalldruckpegel). Wird nur der Mittenkanal codiert und die beiden Seitenkanäle heraus gefiltert, verliert codierte Musik an Qualität. Wird großer Wert auf die Qualität gehaltvoller Musik gelegt, ist vom Einsatz dieser Technik abzuraten.

Eine weitere Möglichkeit, die hier nur kurz erwähnt werden soll, ist die Technik, hochfrequente Subbänder zu einem Band zusammenzufassen und gemeinsam zu quantisieren. Auch hierdurch fällt ein hoher Informationsverlust an, im Positiven (geringerer Datenstrom), wie im Negativen (schlechte Qualität), wobei betont werden sollte, dass der negative Aspekt hier überwiegt. In der Darstellung der Effizienz der Verfahren im Kapitel wird deutlich, dass die Anwendung des Verfahren nur in besonders geeigneten Fällen, wie zum Beispiel bei der Sprachübertragung live übers Internet, wegen der hohen Kompressionsrate berechtigt ist.

4. Praktische Umsetzung in Encodern

4.1 Frequenzbänder

Sogenannte Subband – Coder führen eine Fast – Fourier – Transformation zur Zerlegung des Frequenzspektrums durch. Im Standard MPEG-1 (Layer 2) wird hier eine Zerlegung auf 32 Subbänder durchgeführt. Dabei liegt das „wichtigste“ Signal auf jeweils nur einem Band, das mit einer hohen Quantisierungsauflösung bearbeitet wird, die anderen Bänder können zu diesem Zeitpunkt vernachlässigt werden und mit niedrigerer Auflösung betrachtet werden, was reduzierend auf den Datenstrom wirkt.

Neuere, höher auflösende Verfahren wie AAC verwenden Transformations – Coder mit bis zu 1024 Bändern⁵, um Genauigkeit der Analyse und somit geringere Bitraten zu erreichen.

4.2 Simultane Maskierung

Wie bereits mehrmals erwähnt unterscheiden auch Audio – Encoder periodische bzw. aperiodische Frequenzen. Ein gutes Unterscheidungsmerkmal in der Praxis ist, dass viele Frequenzen, die nahe beieinander liegen (also oft innerhalb eines kritischen Bandes), auf ein tonales Signal hindeuten.

⁴ Aus Quelle[4] genaue WWW – Seite: <http://www.tecchannel.de/multimedia/58/6.html>

⁵ Aus Quelle[3]

Die Berechnung der „Spread – of – Masking“ – Funktion ist mathematisch sehr schwierig. In vielen Fällen wird diese daher näherungsweise durch eine dreieckige Funktion mit den Steigungen +25dB und –10dB (pro Bark) berechnet.

Aperiodische Signale werden bei der Quantisierung zuerst bearbeitet. Sie haben bessere Verdeckungseigenschaften, verdecken deshalb statistisch mehr Signale, deren Quantisierung dann überflüssig wird.

5. Beispiele und Daten

5.1 Beispiel MPEG-1

Der MPEG-1 Standard unterteilt das Verfahren der psychoakustischen Codierung in fünf Schritte:

- (1) Spektrale Analyse und SPL – Normierung
- (2) Klassifizierung des Eingangssignals in tonale und atonale Maskierer
- (3) Reduzierung der Maskierer
- (4) Berechnung der individuellen und der
- (5) globalen Hörschwelle

Das Ziel der spektralen Analyse (1) ist eine hoch aufgelöste Zerlegung des Eingangssignals zu erreichen. „Die SPL – Normierung garantiert das ein 4 kHz – Signal mit einer Amplitude von ± 1 (nach der Normierung) in der Nähe der 0 – Dezibel – Schranke liegt“⁶.

Bei der Klassifizierung (2) in tonale bzw. atonale Maskierer gilt folgende Regel: Liegt innerhalb einer Δ – Umgebung bezogen auf die SPL ein lokales Maximum, das sich deutlich von der Umgebung abhebt⁷ vor, wird dieses als tonal eingestuft, alle anderen Maxima werden als atonale Maskierer aufgefasst. Anschaulich gilt: Liegen viele Signale und ein Maximum mit hoher Differenz zu den benachbarten Signalen (wieder bezogen auf die SPL) innerhalb eines Barks schließen Codierer auf einen tonalen, andernfalls auf einen atonalen Maskierer.

Im nächsten Schritt (3) werden überflüssige Maskierer (Maskierer unterhalb der Mithörschwelle) aus der erstellten Liste der Maskierer gestrichen. Außerdem werden zwei Maskierer, deren Bark – Differenz nur 0.5 Bark beträgt, durch den stärkeren der beiden (bezogen auf die SPL) ersetzt.

Nun wird getrennt für tonale und atonale Anteile eine „individuelle Maskierungsschwelle“⁸ berechnet (4).

Im letzten Schritt (5) werden nun die beiden in (4) berechneten Maskierungsschwellen zusammengesetzt.

⁶ Aus Quelle[1] Seite 462 F. Example Codec ... „Step 1 Spectral Analysis and SPL Normalization“

⁷ Quelle[1] nennt eine Differenz von 7dB, Seite 464

⁸ Quelle[1] Seite 466 (Step 4)

5.2 Daten für ausgewählte Audioverfahren

Um den Kompressionseffekt von MP3 zu verdeutlichen seien zunächst Audio – CD und MP3 gegenübergestellt. Das gesamte Verfahren der MP3 – Codierung, patentiert vom Fraunhofer – Institut, nicht nur die psychoakustische Zerlegung, (siehe Kapitel 5.1) wird in Quelle[11] ausführlich beschrieben. Tabelle 1 zeigt die Gegenüberstellung:

VERFAHREN	BITRATE (kBit/sec.)	KOMPRESSION	EINSATZ
PCM	1114,2	1	„normale“ CD
MPEG-1 Layer III	112 – 128	10 – 12	MP3

Tab. 1: Bitraten

In der folgenden Tabelle werden anhand von Quelle[9] unterschiedliche Codierungsverfahren in ihrer Qualität bewertet:

CODIERER	96 kBit/sec.	128 kBit/sec.	160 kBit/sec.	192 kBit/sec.
MPEG AAC	-1,15	-0,47	-	-
MPEG Layer III	-	-1,73	-	-
Dolby AC-3	-	-2,11	-1,04	-0,52
MPEG Layer II	-	-2,14	-1,75	-1,18

Tab. 2: Qualitätsvergleich der Codierer

Zur der in Tabelle (2) verwendeten Bewertungs – Methode sind folgende Anmerkungen zu machen:

„Ein Differenzgrad von 0 heißt, daß keine Differenzen zu hören sind, -1 heißt hörbar aber nicht gestört, -2 bedeutet, das Material ist leicht gestört, -3 gestört, -4 stark gestört. Die Gesamtraten sind für Stereosignale gezeigt. Der Testfehler für die Differenzgrade liegt bei +/- 0.111.“⁹

5.3 Effizienz der vorgestellten Verfahren

In diesem Kapitel sollen die Kompressionseffekte der unterschiedlichen Phasen bei der MP3 – Codierung dargestellt werden:

ARBEITSSCHRITT	KOMPRESSIONSEFFEKT
Simultane & Zeitliche Maskierung	5 – 6
Stereo – Reduktion	1
Huffman – Encoding	5

Tab. 3: Kompressionseffekte

Das in dieser Arbeit nicht näher beschriebene Huffman – Encoding hat also neben den psychoakustischen Methoden bei der Codierung von Audio – Strömen hohe Kompressionseffekte. Nähere Informationen zum Verfahren des „Huffman – Encodings“ findet der interessierte Leser unter Quelle[11] und Quelle[4].

⁹ Quelle[9]

Bei der in Kapitel 3.4 angesprochenen Technik „hochfrequente Subbänder zusammenzufassen“ werden Datenraten von bis zu 64 kBit/sec. erreicht¹⁰. Ohne hierauf näher einzugehen, sei, wie oben bereits beschrieben, erwähnt, dass dies der Übertragungsrate bei ISDN entspricht. Deshalb ist der Einsatz dieser Methode in manchen Fällen, hierdurch belegt, sinnvoll. Dies zeigt die Effizienz, wie den angesprochenen hohen Qualitätsverlust.

¹⁰ Quelle[10]

Abbildungen - und Quellenverzeichnis

Der Großteil des Vortrags wurde mit Hilfe des Papers Quelle[19] erstellt. Die Angaben im Beispiel 1 wurden Quelle[2] entnommen. Die Quellen zu Kapitel 1.4.1 sind [5], in Kapitel 1.4.2 wurden die Quellen [6], [7] und [8] verwendet. Alle weiteren verwendeten Quellenangaben finden sich im Text.

- [1] Andreas Spanias 17.11.1999 Perceptual Coding of Digital Audio
Proceedings of the IEEE, Vol. 88, No. 4, April 2000
- [2] Hermann Fritz: Wissenswertes zur Schallanalyse
<http://unet.univie.ac.at/~a7425519/Skripten/Schallanalyse.html#I>
- [3] Aus der Internet – Präsenz von "Sound - Studienbereich Neue Medien - Zürich":
<http://snd.snm-hgkz.ch/~alex/SoundII/doc/SoundII-175.htm>
- [4] Internet – Präsenz von „TecChannel“: <http://www.techannel.de/>
- [5] Internet – Präsenz der FH Hannover:
http://www.ik.fh-hannover.de/ik/personen/huthloff/haase/2_1.htm#v9
- [6] Internet – Präsenz von O'Reilly:
„Dictionary of PC Hardware and Data Communications Terms“
http://www.oreilly.com/reference/dictionary/terms/P/Pulse_Code_Modulation.htm
- [7] Internet – Präsenz zu „Speech – Analysis“ von Tony Robinson
<http://svr-www.eng.cam.ac.uk/~ajr/SA95/node81.html>
- [8] Internet – Präsenz zu „Speech – Analysis“ von Tony Robinson
<http://svr-www.eng.cam.ac.uk/~ajr/SA95/node82.html>
- [9] Diplomarbeit von Andy Schimmelpfennig
<http://mitglied.lycos.de/AndiSofti/studies/Mpeg1L3/Auszug.htm#12>
- [10] Markus Conrath: Psychoakustik
http://user.cs.tu-berlin.de/~markusso/down/9812_tub_aes_psychoakustik_folien.pdf
- [11] Internet – Präsenz der FH Jena
http://www.fh-jena.de/contrib/fb/et/personal/ansorg/mp3/mp3_2.htm

Abb. 1: Generisches Modell eines Audio – Encoders aus Quelle [1], Fig. 1.: Generic Perceptual audio encoder

Abb. 2: Beispiel (2): NMT angelehnt an Quelle[1], Fig. 8.(a) : Example to illustrate the asymetrie of simultaneous masking

Abb. 3: Beispiel (3): TMN angelehnt an Quelle[1], Fig. 8.(b) : Example to illustrate the asymetrie of simultaneous masking

Abb. 4: Vor – und Rückwärtsmaskierung aus Quelle[1], Fig. 10. Nonsimultaneous masking properties of the human ear

Tab. 1: Bitraten, Quelle: Eigene Erstellung unter Verwendung der Quellen zu Beispiel 1, und [1]

Tab. 2: Qualitätsvergleich der Codierer Quelle[9]

Tab. 3: Kompressionseffekte, Quelle: Eigene Erstellung unter Verwendung von Quelle [10]