

Audio

Jörg Cassens

Institut für Mathematik und Angewandte Informatik

Medieninformatik
WS 2017/2018




medieninformatik

IMAI – Institut für
Mathematik und
Angewandte Informatik

Pingo

Schall
Wahrnehmung
Digitalisierung
Kompression
Bearbeitung
Literatur



 pingo.upb.de/132702

Einleitung

- Visuelle Reize vielleicht dominant, Töne aber sehr wichtig
 - Wir können wegschauen, aber nicht weghören
- Einzig steuerbarer Parameter häufig die Lautstärke
- Akustische Warnsignale deshalb akustisch weil wirksam
- Wahrnehmung oft unbewußt, ungewollt, nebenher
- Beeinflussung der Wahrnehmung
 - Filmmusik
- Transportiert z.B. Stimmungen anderer Personen
- Definition von Kulturen über die Musik
- Gesellschaftliche Funktion
 - Hymnen
 - Kampflieder
 - Stadiongesänge
- Audiodesign inzwischen wesentlicher Bestandteil der Produktentwicklung

Schall

Wahrnehmung

Digitalisierung

Kompression

Bearbeitung

Literatur

- Physikalische Grundlagen akustischer Signale
 - Schallwellen
 - Ausbreitung
- Wahrnehmung
- Aufnahme
- Digitalisierung
- Kompression
- Audiotbearbeitung

Outline

Schall

Wahrnehmung

Digitalisierung

Kompression

Bearbeitung

Literatur

1 Schall

2 Wahrnehmung

3 Digitalisierung

4 Kompression

5 Bearbeitung

- Mechanische Bewegung eines physikalischen Mediums wie Luft oder Wasser
 - Explosionen im Weltall
- Breitet sich als Druckwelle aus
- Unterschiedliche Ausbreitungsgeschwindigkeit in unterschiedlichen Medien (dichter ist schneller)
- In Luft etwa 331 m/s bei 0 Grad Celsius, 343 m/s bei 20 Grad
- Zumeist periodische Signale
- Vom Menschen wahrnehmbar etwa 18 Hz bis 20 kHz (etwa 17 m bis 1,7 cm)

Schall

Wahrnehmung

Digitalisierung

Kompression

Bearbeitung

Literatur

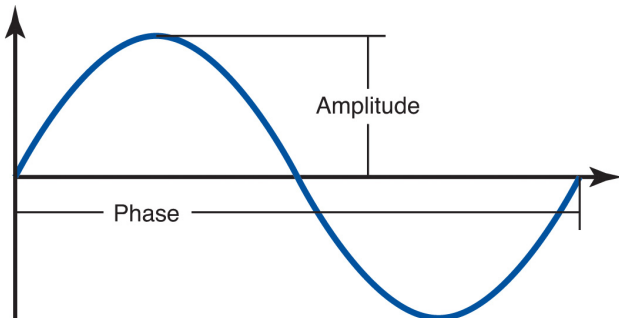


Abbildung 4.1: Grundlegende Größen eines periodischen Signals

Schall

Wahrnehmung

Digitalisierung

Kompression

Bearbeitung

Literatur

- **Amplitude:** Maximaler Wert in einer Schwingung
- **Phase:** Ein Durchlauf
- **Phasenverschiebung:** Verschiebung des Nulldurchgangs (0-360 Grad)
- **Frequenz:** Anzahl der Schwingungen pro Zeiteinheit
- **Wellenlänge:** Der Weg, den das Signal bei gegebener Ausbreitungsgeschwindigkeit zurücklegt

Longitudinalwelle vs. Transversalwelle

Schall

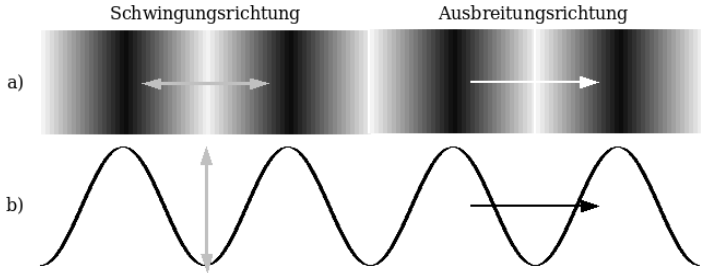
Wahrnehmung


Digitalisierung

Kompression

Bearbeitung

Literatur



 Wikipedia – User Debianux

Longitudinalwelle

Schall

Wahrnehmung

Digitalisierung

Kompression

Bearbeitung

Literatur

 [Wikipedia](#)

Hörbare Frequenzen

Schall

Wahrnehmung

Digitalisierung

Kompression

Bearbeitung

Literatur

- Der vom Menschen hörbare Frequenzraum
- Beachte: für Sprache reicht ein weitaus kleinerer Teil aus

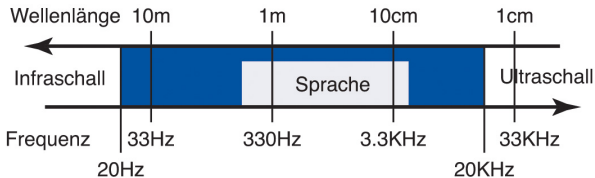


Abbildung 4.2: Hörbarer Frequenzbereich: Sprache nutzt nur einen Teil der hörbaren Frequenzen aus.

- Konzentrische Ausbreitung um den Entstehungspunkt
- Intensität nimmt mit dem Quadrat der Entfernung ab
- Schallwellen können absorbiert, reflektiert, gebrochen und gebeugt werden
- Schallwellen werden an Objekten gebeugt, deren Abmessungen in etwa in der Größenordnung der Wellenlänge liegen
- Liegt in Bereichen, bei denen Elemente der Architektur eine Rolle spielen können
 - Tiefe Frequenzen “fließen” um eine Säule herum
 - Hohe Frequenzen werden abgeblockt

- Wesentliche Aspekte lassen sich z.B. bei der Betrachtung von Saiteninstrumenten erkennen
- Wird eine Saite in Schwingungen versetzt entstehen mehrere Töne
- Eine Saite schwingt zwischen den festen Elementen
- Mit relativ wenig Energiezufuhr kommen Frequenzen aus, die an den Fixpunkten einen Nulldurchgang haben
 - Grund- oder Resonanzfrequenz
 - Vielfache dieser Schwingung (Obertöne)
- Übertragbar auf andere Klangerzeuger
 - Flächen (Trommeln)
 - Luftsäulen in den Blasinstrumenten

■ Harmonische oder Obertöne

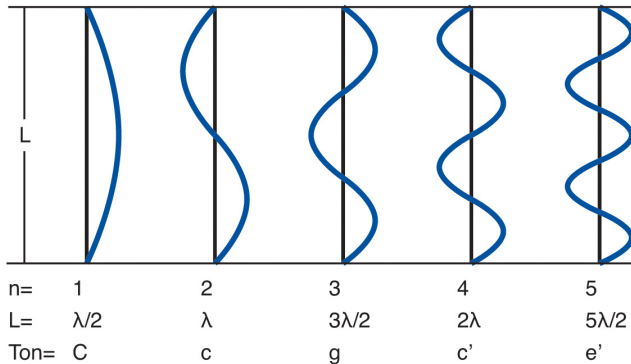


Abbildung 4.3: Oberwellen sind ganzzahlige Vielfache (n) einer Grundfrequenz.

Obertöne: Beispiel

Schall

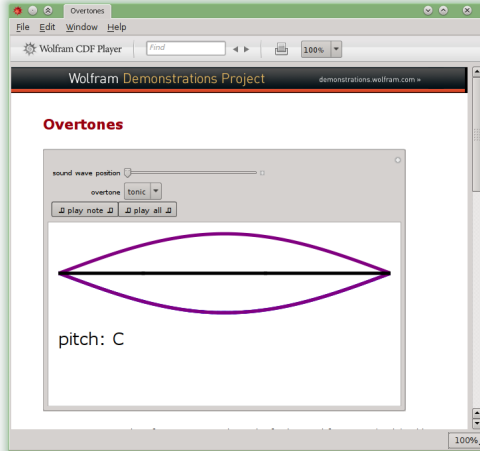
Wahrnehmung


Digitalisierung

Kompression

Bearbeitung

Literatur



 demonstrations.wolfram.com/Overtones

Tonleiter

- Westliches Tonsystem mit den zwölf Tönen einer Oktave läßt sich näherungsweise aus Obertönen konstruieren
- Sei f die Frequenz des Grundtons, dann hat die Oktave darüber die doppelte Frequenz, $2f$, die reine Quinte über der Oktave $3f$, die große Terz $5f$ etc.
- Obertonreihen erzeugen *reine* Stimmung
- In der westlichen musikalischen Praxis werden *temperierte* Stimmungen verwendet (gleich klingende Intervalle) bezüglich der Grundtöne
 - Gezielt etwas unrein gestimmt, um kleine Tohnhöhenunterschiede zu verteilen
 - Nicht-Übereinstimmung von Quinten und Terzen
 - Verschiedene Tonarten mit 12 Tasten
 - Zusammenspiel unterschiedlicher Instrumente
- J.S. Bach: Das wohltemperierte Klavier
- Blechblasinstrumente spielen in reiner Stimmung

■ Analog Histogramm bei Bildern

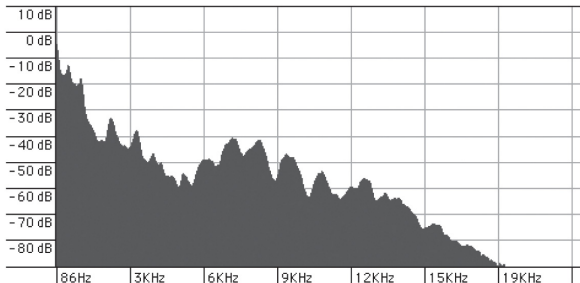


Abbildung 4.5: Im Frequenzspektrum sind die in einem Signal enthaltenen Frequenzen auf der horizontalen Achse und ihre jeweiligen Intensitäten auf der vertikalen Achse aufgetragen.

- Bei Mischung von 2 Signalen werden diese addiert
- Mischung mit sich selber: konstruktive Interferenz
- Mischung mit invertiertem (180 Grad phasenverschobenem) Signal – destruktive Interferenz
- Mischung zweier nahe beieinander liegenden Signale kann zur Schwebung führen
 - Nutzbar bei der Stimmung von Instrumenten
- Bei der Mischung zweier Signal kann das eine das andere modulieren
 - Frequenz, Amplitude, Phase
 - Grundlage erster Synthesizer
- Nicht-periodische Signale (Geräusch)

Addition

Schall

Wahrnehmung

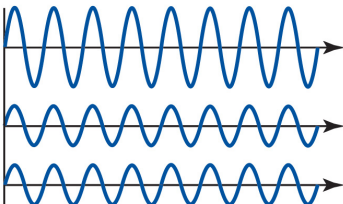
Digitalisierung

Kompression

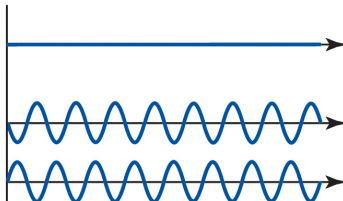
Bearbeitung

Literatur

Konstruktive Interferenz



Destruktive Interferenz



Schwebung

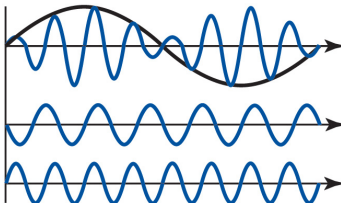


Abbildung 4.6: Konstruktive Interferenz, destruktive Interferenz und Schwebung (abwechselnd konstruktive und destruktive Interferenz)

Addition: Beispiel

Schall

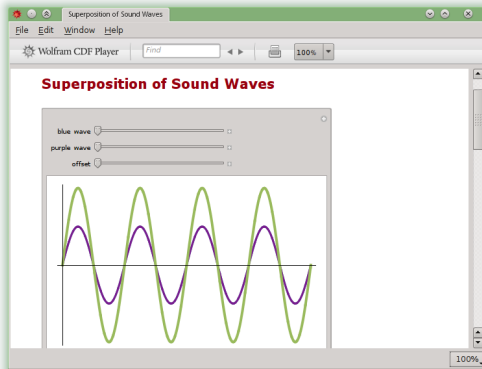
Wahrnehmung

Digitalisierung

Kompression

Bearbeitung

Literatur



☞ demonstrations.wolfram.com/SuperpositionOfSoundWaves

Annäherung: Rechteck

Ungerade Vielfache der Grundfrequenz

Schall

Wahrnehmung

Digitalisierung

Kompression

Bearbeitung

Literatur

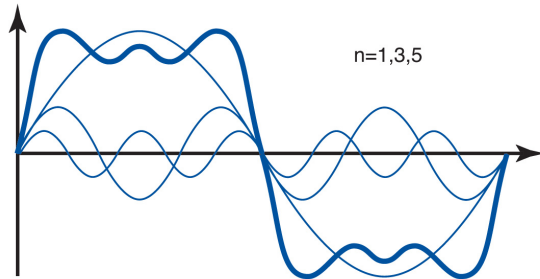
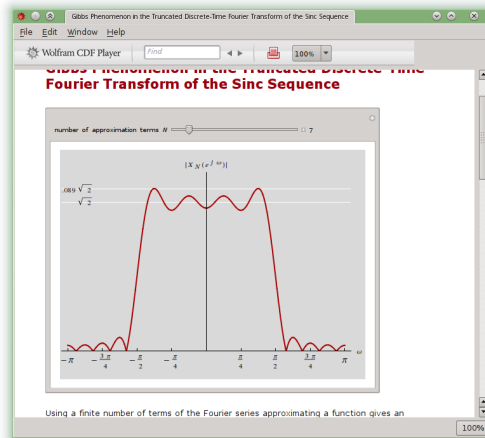


Abbildung 4.4: Annäherung eines Rechtecksignals durch ungerade Vielfache der Grundfrequenz. Durch Addition weiterer ungerader Vielfacher kann die Signalform beliebig gut angenähert werden.



[☞ demonstrations.wolfram.com/](https://demonstrations.wolfram.com/)

GibbsPhenomenonInTheTruncatedDiscreteTimeFourierTransformOf/

Outline

Schall

Wahrnehmung

Digitalisierung

Kompression

Bearbeitung

Literatur

1 Schall

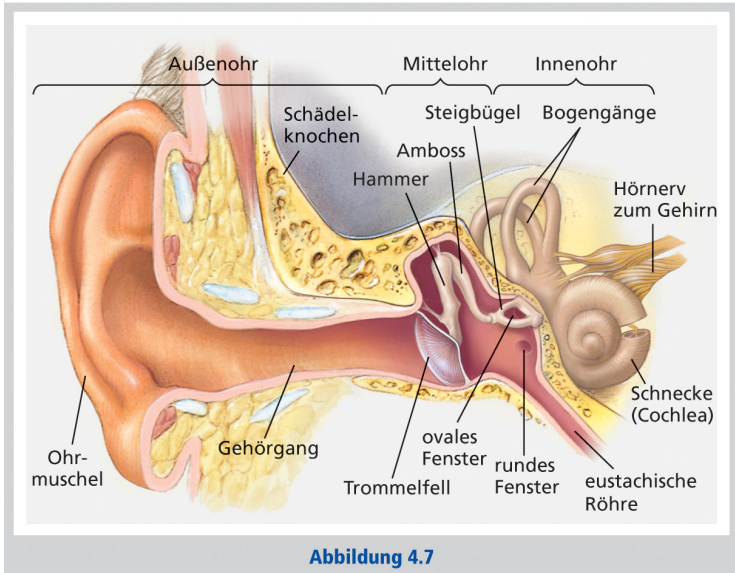
2 Wahrnehmung

3 Digitalisierung

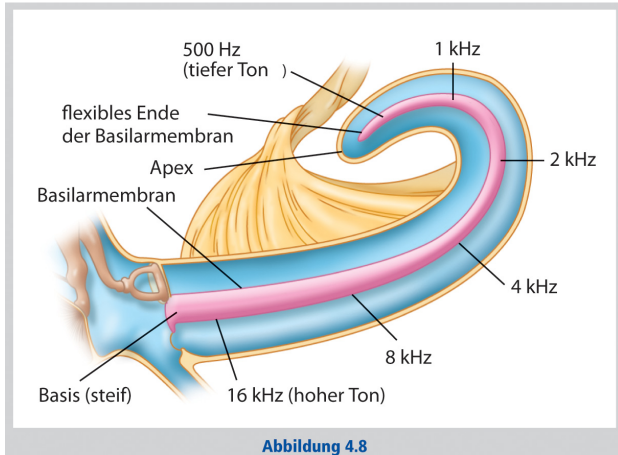
4 Kompression

5 Bearbeitung

Akustische Wahrnehmung



Akustische Wahrnehmung (contd.)



Maskierung

- Bei nahe beieinander liegenden Reizen kann ein energiereiches Signal eine nahe liegendes, schwächeres Signal maskieren (Frequenzmaskierung)
- Auch schwache Reize vor oder nach starken Reizen (zeitliche Maskierung)

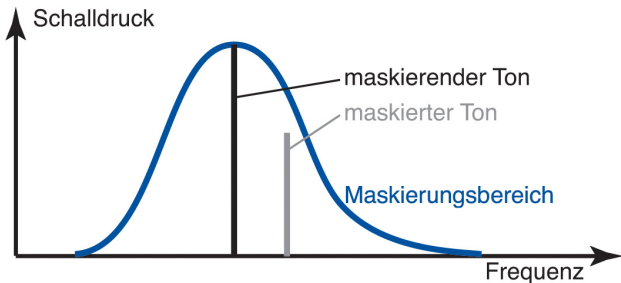


Abbildung 4.9: Maskierung eines leisen Tones durch einen benachbarten lauten Ton

Maskierung: Beispiel

Schall

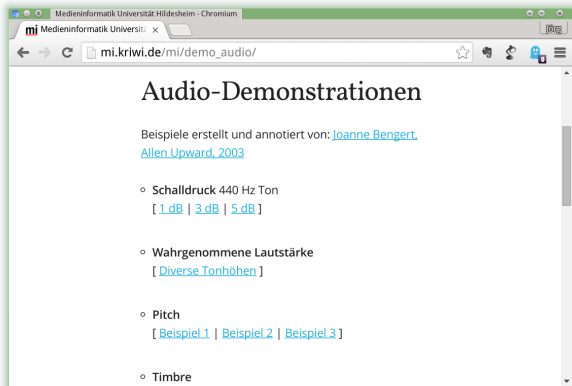
Wahrnehmung

Digitalisierung

Kompression

Bearbeitung

Literatur



🔊 Audio Demo-Dateien

- Messung in Dezibel (dB)
- Logarithmische Skala (6 dB) Verdoppelung des Schalldrucks
- Verdoppelung der Lautstärke? It's complicated...
 - Physischer Schalldruckpegel stark abhängig von Frequenz
 - Unterschiede bei ansteigender Lautstärke stärker wahrgenommen als bei abnehmender
 - Gewöhnungseffekte z.B. bei lauter Musik, dieser Effekt ist aber nur kurzzeitig wirksam
 - Geht in die Psychoakustik
- Hörschwelle: leiseste bei 2 kHz gerade noch hörbare Geräusch = 0 dB
- Charakteristische Lautstärke
 - Sprache: 30-70 dB
 - Schmerzwahrnehmung: 120 dB

Lautstärke: Beispiel

Schall

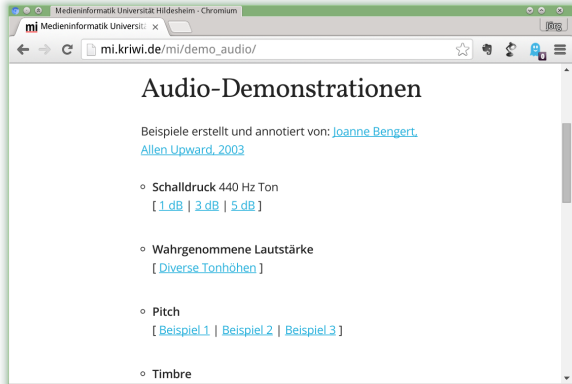
Wahrnehmung

Digitalisierung

Kompression

Bearbeitung

Literatur



Medieninformatik Universität Hildesheim - Chromium

mi Medieninformatik Universität x

mi.kriwi.de/mi/demo_audio/

Audio-Demonstrationen

Beispiele erstellt und annotiert von: [Joanne Bengert](#),
[Allen Upward, 2003](#)

- Schalldruck 440 Hz Ton
[[1 dB](#) | [3 dB](#) | [5 dB](#)]
- Wahrgenommene Lautstärke
[[Diverse Tonhöhen](#)]
- Pitch
[[Beispiel.1](#) | [Beispiel.2](#) | [Beispiel.3](#)]
- Timbre

🔊 Audio Demo-Dateien

Lautstärke (contd.)

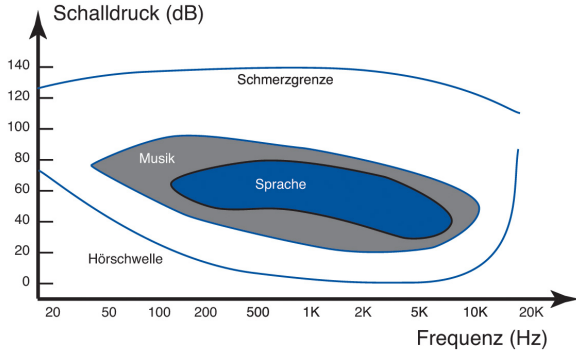
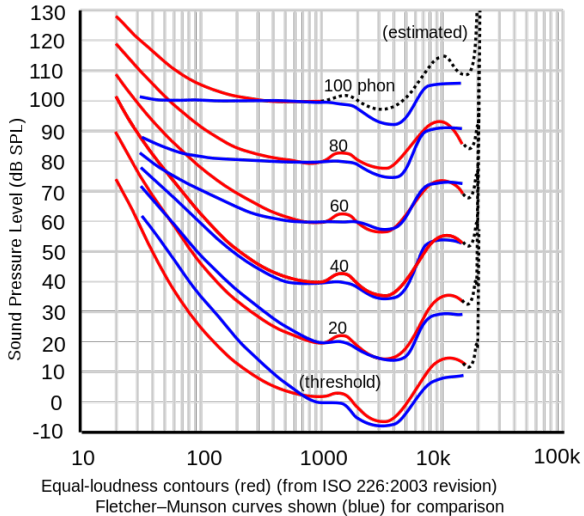


Abbildung 4.10: Lautstärkebereiche des menschlichen Hörens in Abhängigkeit von der Frequenz (vereinfacht)

Lautstärke und Frequenz



Wikipedia

Schall

Wahrnehmung

Digitalisierung

Kompression

Bearbeitung

Literatur

- **Interaurale Zeitdifferenz (ITD):** Unterschied der Laufzeit zwischen beiden Ohren. Gleiche zeitliche Unterschiede führen bei höheren Frequenzen zu höheren Phasenverschiebungen, diese sind besser wahrzunehmen
- **Interaurale Intensitätsdifferenz (IID):** Teilweise Abschwächung der Lautstärke, vor allem bei hohen Frequenzen
- **Head-Related Transfer Function (HRTF):** bei dem Weg um den Kopf werden bestimmte Frequenzanteile verstärkt, andere verschluckt

Richtungshören (contd.)

Schall

Wahrnehmung

Digitalisierung

Kompression

Bearbeitung

Literatur

- Wir versuchen, zusammen mit ITD, IID und HRTF die Richtung zu bestimmen
 - besser bei hohen Frequenzen
- ITD und IID würden für vor unter hinter uns liegende Geräuschquellen das gleiche Ergebnis bringen
- Allein HRTF unterscheidet diese
- Auch: Einbringung von Wissen
 - Flugzeuge sind über uns
- Hilfe durch andere Sinne
 - Die See ist vor uns

Richtungshören (contd.)

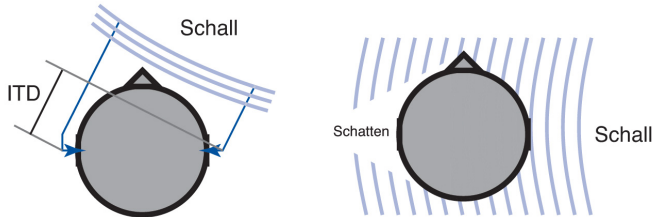


Abbildung 4.11: Der Weg des Schalls zum abgewandten Ohr ist weiter. Daher kommt der Schall dort mit einer zeitlichen Verzögerung (ITD) an. Hohe Frequenzen werden auf der abgewandten Seite abgeschattet, während tiefe Frequenzen am Kopf gebeugt werden und daher weniger abgeschwächt werden.

Outline

Schall

Wahrnehmung

Digitalisierung

Kompression

Bearbeitung

Literatur

1 Schall

2 Wahrnehmung

3 Digitalisierung

4 Kompression

5 Bearbeitung

- Was sind geeignete Werte für Diskretisierung und Quantisierung?
- Vor der Digitalisierung beschränken auf etwa 20-20.000 Hz
 - Abtasttheorem sagt aus, wie oft wir messen müssen
- Wahrnehmbare Lautstärke etwa zwischen 0 dB und 120 dB (Schmerzgrenze)
 - Dynamikumfang etwa 120 dB
 - Da 6 dB einer Verdoppelung der Signalstärke entsprechen ist der Dynamikumfang etwa $2^{120/6} = 2^{20} : 1$
 - Etwa 20 bits reichen aus um den vollen Umfang der menschlichen Lautstärkewahrnehmung auszudrücken
- Caveat: Was sind die kleinsten wahrnehmbaren Lautstärkeunterschiede?
 - Daumenregel 1 dB
- In der Praxis 16 (CD) bis 24 Bit

Abtasttheorem: Beispiel

Schall

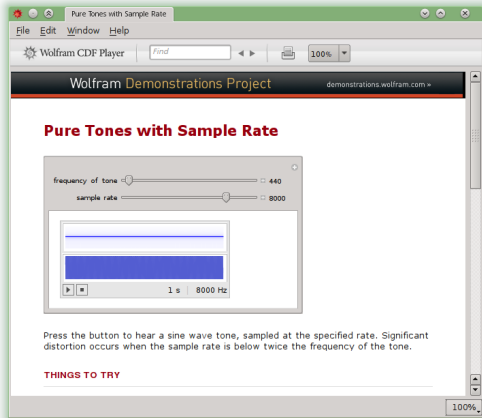
Wahrnehmung

Digitalisierung

Kompression

Bearbeitung

Literatur



 demonstrations.wolfram.com/PureTonesWithSampleRate

- Pulse Code Modulation
- Samples als Bitfolgen liegen hintereinander
- Unkomprimiertes Format
- Standard G.711 für die Sprachübertragung
 - Abtastfrequenz 8 kHz
 - Auflösung 8 Bit
 - ISDN B-Kanal mit 64 kBit/s
- Differential PCM: nur die Differenz zwischen Samples übertragen
- Adaptive Differential PCM: Auch Anzahl Bit variabel

Outline

Schall

Wahrnehmung

Digitalisierung

Kompression

Bearbeitung

Literatur

1 Schall

2 Wahrnehmung

3 Digitalisierung

4 Kompression

5 Bearbeitung

- MPEG = Moving Picture Expert Group
- Gremium der ISO und der IEC (International Electrotechnical Commission)
- MPEG-1 und MPEG-2 für Video mit integriertem Audio
- Codierungen für Bild und Ton jeweils getrennt beschrieben
- Außerdem: Varianten, sogenannte Layer
- MP3 ist MPEG-1 Layer 3
- MPEG auch verwendet bei
 - DAB (Digital Audio Broadcast)
 - DVB (Digital Video Broadcast)
 - DVD

Psychoakustik

- Verlustfreie Verfahren führen zu relativ wenig Kompression
- MP3 basiert auf einem psychoakustischen Modell der Tonwahrnehmung
- Nutzt Maskierungseffekte aus
- Empirisch ermittelt
- 27 so genannte kritische Bänder
 - Ein kritische Band ist der Bereich, in dem Töne stärker als ein bestimmter Grenzwert miteinander interagieren
 - Eine solche Interaktion ist z.B. die Maskierung

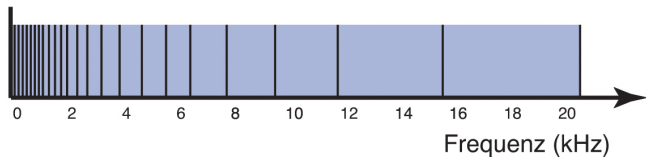


Abbildung 4.12: Kritische Bänder des psychoakustischen Modells

- **Musikalische Betrachtung**
 - Kritische Bänder sind in der Nähe des mittleren C eines Klaviers etwa eine Terz breit
 - Zu den hohen Tönen enger
 - Zu den tiefen Tönen weiter
- **Frequenzbetrachtung**
 - Auf linearer Skala
 - Hohe Frequenzen breiter
 - Tiefe Frequenzen schmaler

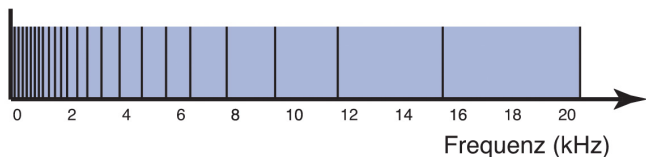


Abbildung 4.12: Kritische Bänder des psychoakustischen Modells

Kritische Bänder

- Aussagen, welche Frequenzen andere Frequenzen maskieren

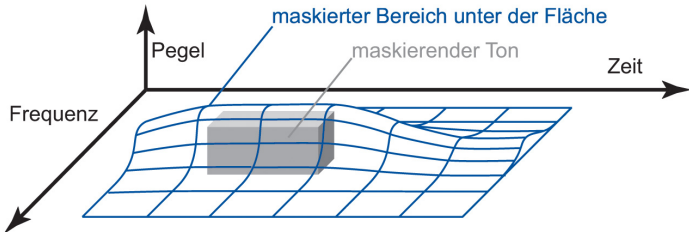
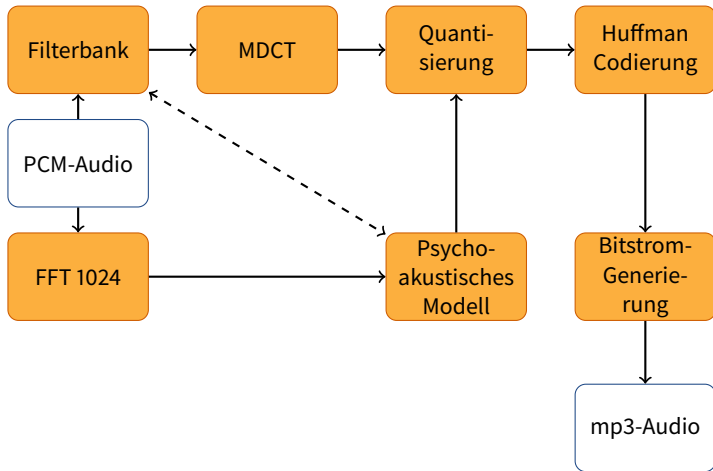


Abbildung 4.13: Frequenzbereich, der durch einen einzelnen Ton räumlich und zeitlich maskiert wird

- Zentrale Idee: Maskierte und Unhörbare Anteile nicht übertragen
- Arbeit im Frequenzraum

- Der Standard beschreibt die Codierung relativ abstrakt
 - Fraunhofer Referenzimplementierung
 - Verschieden Encoder produzieren unterschiedliche Ergebnisse
- Einige Stufen werden definiert
 - Bitraten 32-320 kbit/s (MPEG-1)
 - Sampling 32-48 kHz (MPEG-1)
- Das Decodieren ist genauer geregelt
 - Unterschiedliche Encoder sollten auf dem gleichen Eingabedaten den gleichen Bitstrom erzeugen (modulo Rundungsfehler)
- Die Beschreibung des Verfahrens ist vereinfacht

Prinzipschaltbild



Vereinfachtes Prinzipschaltbild nach Watkinson [2004]

Sampling

- Sehr kurze Zeitabschnitte (36 samples = 0,8 ms)
- Mit Hilfe einer Filterbank in 32 Bänder zerlegt
- “Windowing”: Auswahl von Zeitabschnitten (Frames) für die weitere Codierung
- Unterschiedliche Arten von “Windows” für unterschiedliche Signale
 - “Long Window” bei wenigen Änderungen
 - “Short Window” bei größeren Änderungen
- Danach Modifizierte Diskrete Kosinustransformation (MDCT)
- Darstellung bzgl. 18 Grundfrequenzen innerhalb der jeweiligen Subbänder
- erreichte Darstellung: 32 (Bänder) * 18 Frequenzen = 576 Basisfrequenzen
- Hohe zeitliche Auflösung

- Parallel wird das Eingangssignal mittels schneller Fourier-Transformation in eine Frequenzraumdarstellung bzgl. 1024 Bändern überführt
- Wesentlich niedrigere zeitliche Auflösung, aber bessere Information über Frequenz und Phase
- Dies ist günstiger für die Berechnung der Maskierung
- Auf Basis dieser Zerlegung wird die Maskierung zwischen einzelnen Frequenzen bestimmt
- Aus den Koeffizienten der FFT und dem psychoakustischen Modell wird berechnet, welche Frequenzanteile des zerlegten Signals maskiert und somit nicht wahrgenommen wird

Verlustbehaftete Codierung

Schall

Wahrnehmung

Digitalisierung

Kompression

Bearbeitung

Literatur

- Genauer: aus den Koeffizienten der FFT und dem psychoakustischen Modell wird berechnet,
 - wie die Quantisierung der Koeffizienten aussieht und
 - welche Window-Größen angewandt werden sollen
- Davon abhängig werden die MDCT-Koeffizienten quantifiziert, wobei viele Nullwerte entstehen
- Die Quantisierung ist auch davon abhängig, welche Bitrate zur Verfügung steht
- Der entstehende Teil wird mit Huffman kodiert
 - Anpassung der Quantisierung an die zur Verfügung stehenden Codetabellen
- Zusätzlich Ausnutzen der Redundanz zwischen Stereokanälen

Prinzipschaltbild Revisited

Schall

Wahrnehmung

Digitalisierung

Kompression

Bearbeitung

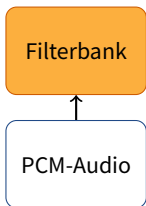
Literatur

PCM-Audio

mp3-Audio

Vereinfachtes Prinzipschaltbild nach Watkinson [2004]

Prinzipschaltbild Revisited

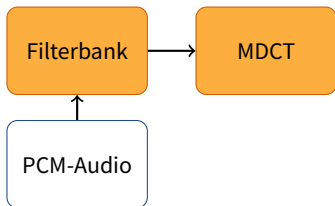


Zerlegung des Eingangssignals in 32 (kritische)
Frequenzbänder und Windowing

mp3-Audio

Vereinfachtes Prinzipschaltbild nach Watkinson [2004]

Prinzipschaltbild Revisited

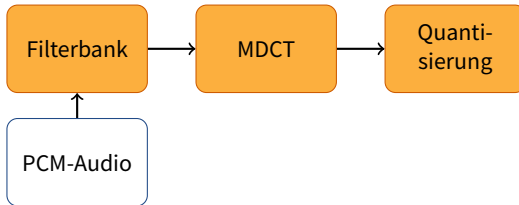


Modifizierte Diskrete Kosinustransformation: Je 18 Grundfrequenzen in jedem der 32 Bänder entsprechend 576 Basisfrequenzen

mp3-Audio

Vereinfachtes Prinzipschaltbild nach Watkinson [2004]

Prinzipschaltbild Revisited

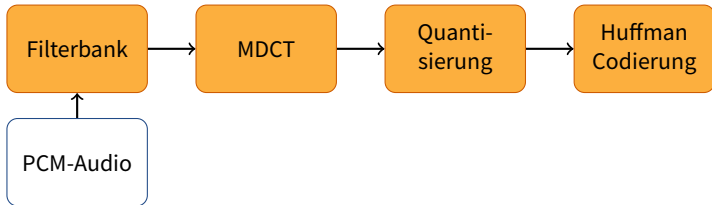


Anpassung der Koeffizienten der MDCT, z.B. auf Null setzen unhörbarer Frequenzanteile und Anpassung an Bandbreite

mp3-Audio

Vereinfachtes Prinzipschaltbild nach Watkinson [2004]

Prinzipschaltbild Revisited

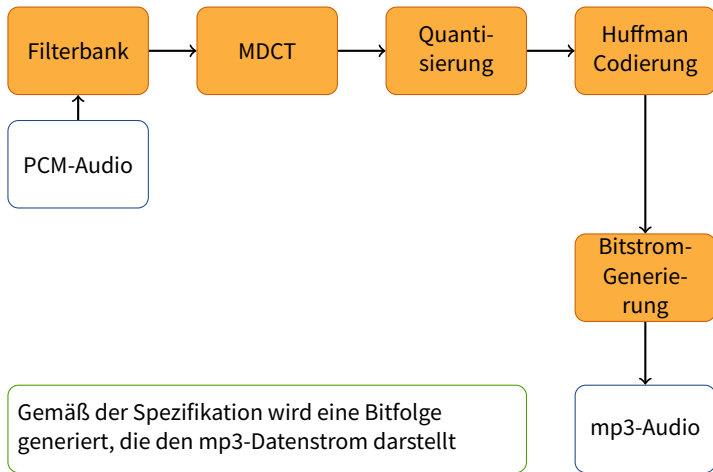


Huffman-Kompression da der Nachrichtenstrom eine hohe Entropie hat (z.B. viele Nullelemente)

mp3-Audio

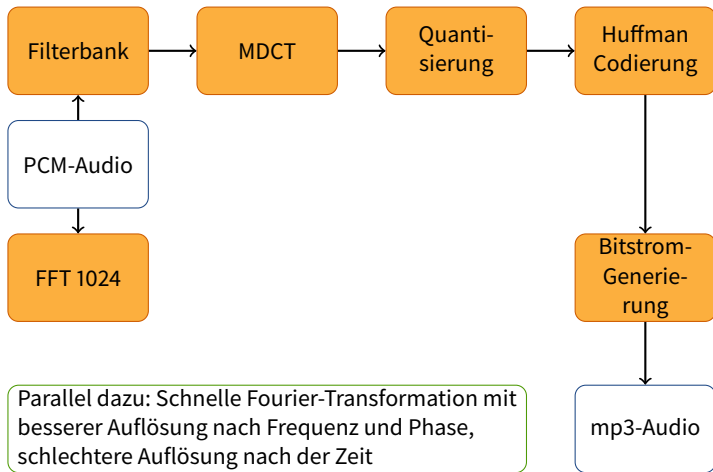
Vereinfachtes Prinzipschaltbild nach Watkinson [2004]

Prinzipschaltbild Revisited



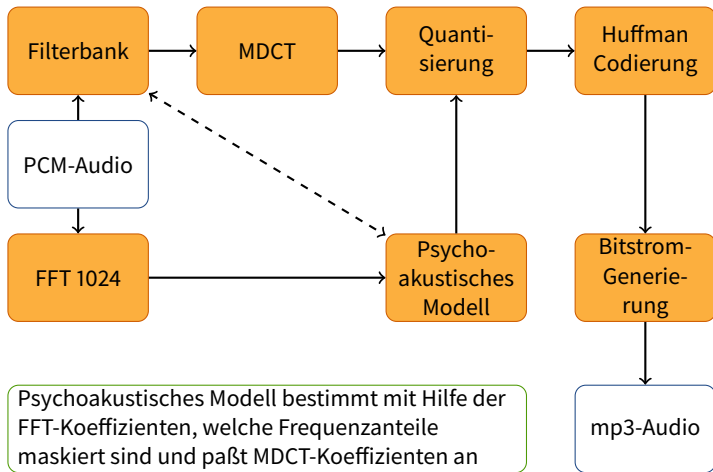
Vereinfachtes Prinzipschaltbild nach Watkinson [2004]

Prinzipschaltbild Revisited



Vereinfachtes Prinzipschaltbild nach Watkinson [2004]

Prinzipschaltbild Revisited



Vereinfachtes Prinzipschaltbild nach Watkinson [2004]

Varianten und Weiterentwicklungen

Schall

Wahrnehmung

Digitalisierung

Kompression

Bearbeitung

Literatur

- Das MP3-Format hat wesentlich dazu beigetragen, daß Musik online verteilt werden konnte
 - Wesentlich reduzierte Dateigrößen bei akzeptabler Qualität
 - Nutzen eines psychoakustischen Modells wurde gezeigt
 - Unabhängige Implementierungen, zum Teil als Freie Software
- Neben MP3 gibt es inzwischen eine Reihe verlustbehafteter Audioformate
 - Advanced Audio Coding (AAC)
 - Ogg/Vorbis
 - Opus
- Auch bei MP3 weitere Optimierungen
 - Variable Bitrate (VBR), durchschnittliche Bitrate (ABR), konstante Bitrate (CBR)
 - Optimierung des psychoakustischen Modells bei der Codierung

Schall

Wahrnehmung


Digitalisierung

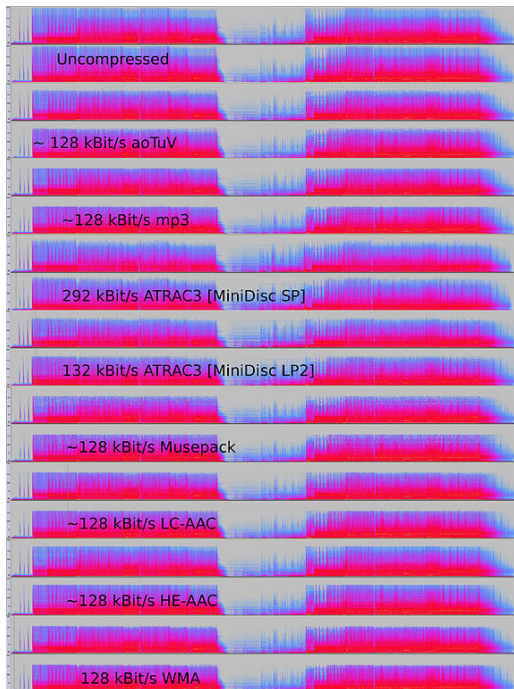
Kompression

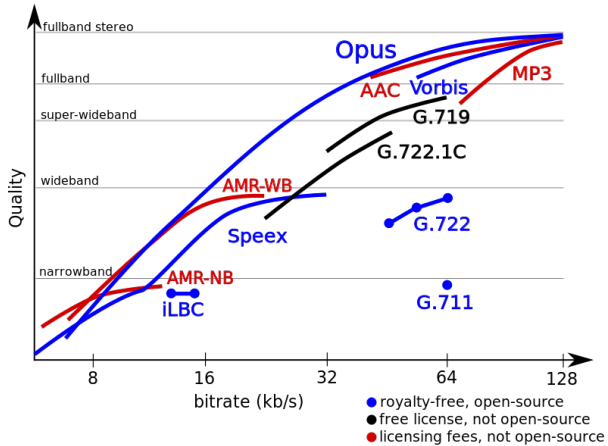
Bearbeitung

Literatur



 Wikipedia





Wikipedia

Verlustfrei

Schall

Wahrnehmung

Digitalisierung

Kompression

Bearbeitung

Literatur

- mp3, Opus, AAC erreichen die hohen Kompressionsraten dadurch, daß sie Verlustbehaftet komprimieren
- Die direkte Speicherung der Samples (PCM) ist sehr speicherplatzintensiv
- Ist eine verlustfreie Speicherung möglich, die effektiver in Hinblick auf den Speicherplatz ist?

- mp3, Opus, AAC erreichen die hohen Kompressionsraten dadurch, daß sie Verlustbehaftet komprimieren
- Die direkte Speicherung der Samples (PCM) ist sehr speicherplatzintensiv
- Ist eine verlustfreie Speicherung möglich, die effektiver in Hinblick auf den Speicherplatz ist?
- Idee: Kombination der Vorteile von verlustbehaftet und verlustfrei
 - Speichere eine verlustbehaftetes Signal mit Hilfe eines effektiven Modells
 - Merke die Unterschiede zum verlustfreien Signal

- **FLAC: Free Losless Audio Codec**
- **Kompression darf rechenintensiv sein, die Dekompression muß in Echtzeit ablaufen können**
- **Beschränkung auf Fixpoint-Berechnungen erlaubt Realisierung in Hardware**
- **Unterstützt**
 - **PCM-Quantisierung von 4-32 Bit**
 - **Sampling-Raten von 1Hz bis etwa 655 kHz**
 - **1-8 Kanäle**
- **FLAC-Datei enthält Metadaten und Audiodaten**
- **Daumenregel: während universelle, verlustfreie Kompressionsverfahren bei Audiodaten etwa 10-20% Kompression erreichen können liegt FLAC bei 40-60%**

FLAC: Vorgehen

- 1 **Zerlegung:** Bilden von Blöcken von 1000-6000 Samples
- 2 **Dekorrelierung:** Ausnutzen von Redundanz zwischen Kanälen (rechts/links bei Stereo, weitere Kanäle)
- 3 **Modellierung:** Predictor für das Audiosignal wird erstellt
 - Verbatim (Kopie)
 - Linear
 - Polynomfunktion oder
 - Linear Predictive Coding
- 4 **Fehlersignal:** Der Unterschied des Predictors zum tatsächlichen Signal (Residual Signal) wird codiert
 - Entropiebasiert, Rice-Kodierung
- 5 **Framing:** Alle Kanäle eines Blocks werden mit Metadaten (Fehlererkennung, Timestamps, ...) in einen Frame geladen

■ **Verbatim**

- Das Audiosignal wird ohne Kompression kopiert

■ **Linear**

- Das Signal wird mit einer Linearfunktion angenähert
- Stille, konstante Töne

■ **Polynomfunktion**

- Das Signal wird mittels einer Polynomfunktionen angenähert
- Schnell, weniger genau als LPC

■ **Linear Predictive Coding**

- Vereinfachtes Modell des Klangerzeugers, z.B. des menschlichen Stimmtraktes
 - Summer am Ende eines Rohres
 - Dazu Zisch- und Knalllaute
- Danach müssen nur die Parameter dieses Modells bestimmt (und übertragen) werden

FLAC: Fehlersignal

- Das Fehlersignal wird mittels eines Entropiebasierten Kompressionsverfahren gespeichert
- Im Gegensatz zu Huffman oder dem Arithmetischen Encoding kommt ein Verfahren zum Einsatz, welches Annahmen über die Wahrscheinlichkeitsverteilung macht
- Rice-Encoding: Codiere kleine Werte mit möglichst kurzen Worten
- Verschiedene Parametrisierung möglich
- Grundsätzlich gut geeignet für geometrische Verteilungen
 - Hier könne Kompressionsraten im Bereich der bei Huffman möglichen Raten erreicht werden
- Qualität des Predictors ausschlaggebend für die Verteilung der Werte des Fehlersignals

Outline

Schall

Wahrnehmung

Digitalisierung

Kompression

Bearbeitung

Literatur

1 Schall

2 Wahrnehmung

3 Digitalisierung

4 Kompression

5 Bearbeitung

- Neben der Audiotbearbeitung auch die Audioverarbeitung
- Tendentielle Unterteilung
 - Verarbeitung: Automatische Verarbeitungsketten
 - Bearbeitung: Interaktive Veränderung von Audiodaten
- Ändern der Amplitude (Lautstärke)
- Wenn benachbarte Signalwerte einbezogen werden, spricht man von Filtern
- Weiterhin kann es sinnvoll sein, Frequenz und Phase zu ändern
- Simulation von Effekten wie Echo oder Hall

- Bezugspegel von 0 dB
- Historisch aus der analogen Audiotechnik:
 - 0 dB = 1 mW, 600 Ohm, Spannung 0.7775 V
- Effektivwert:

$$s_{\text{effektiv}} = \sqrt{\frac{1}{T} \int_T s^2(t) dt}$$

- Maximalpegel: Höchster darstellbarer Signalwert
- Arbeitspegel sollte genug Abstand halten

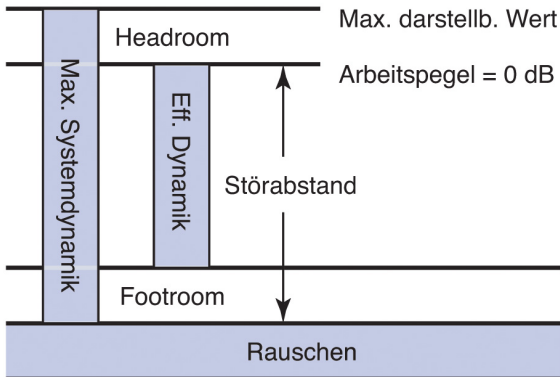


Abbildung 4.15: Headroom, Footroom, Dynamik und Störabstand eines Signals

Pegelanpassung

Schall

Wahrnehmung

Digitalisierung

Kompression

Bearbeitung

Literatur

- **Übersteuerung:** höchste Signalwerte liegen außerhalb des darstellbaren Bereichs der Quantisierung
 - An den Rändern Clipping; starke Änderungen im Signalverlauf (Unstetigkeiten)
 - Wirken sich als hohe Frequenzanteile aus
 - In analoger Technik nicht so stark ausgeprägt, da der Abfall i.d.R. nicht so steil ist
- **Untersteuerung:** Nutzt nicht den gesamten Darstellungsbereich der Quantisierung aus
 - Höchste Signalwerte liegen deutlich unter dem größten darstellbaren Wert
 - Störungen mit gleicher Amplitude sind also im Verhältnis lauter
 - Signal-Rausch-Abstand schmilzt
- **Abhilfe:** Normalisierung
- **Ebenso** Kompensation DC-Offset

Veränderung Hüllkurve

- z.B. Ein-/Ausblenden

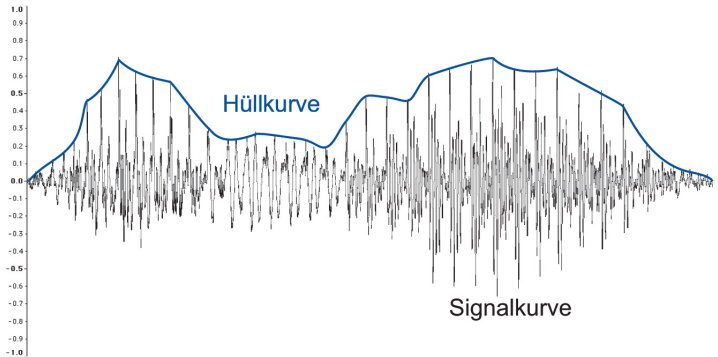


Abbildung 4.16: Hüllkurve und Signalkurve eines Tonsignals

- Benachbarte Werte sind *zeitlich* benachbart
 - Bilder: *räumlich* benachbart
- Analogtechnik: Verwendung von Bauteilen, die ein zeitabhängiges Verhalten aufweisen
 - Spulen
 - Kondensatoren
- Digitaltechnik: Algorithmisch, z.B. Digital Signal Processor (DSP)

Frequenzfilter

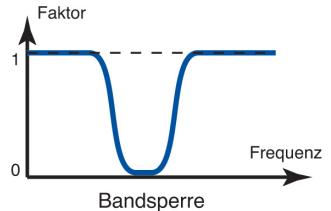
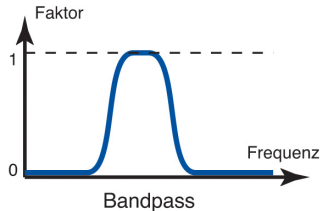
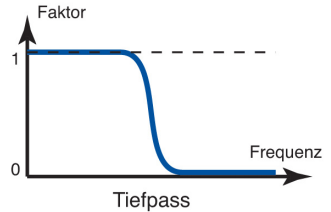
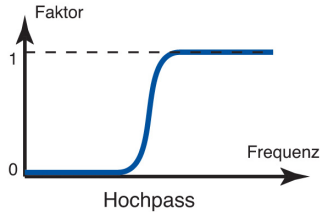


Abbildung 4.17: Die vier grundlegenden Arten von Frequenzfiltern

- Analog zu den Faltungsfilters der Bildbearbeitung
- Weichzeichner und Tiefpaß (Boxcar)

$$g(x) = \frac{1}{3}(1, 1, 1) * (f(x-1), f(x), f(x+1))$$

- Hochpaß

$$g(x) = (-1, 3, -1) * (f(x-1), f(x), f(x+1))$$

- Parametrisch oder (hier) grafisch

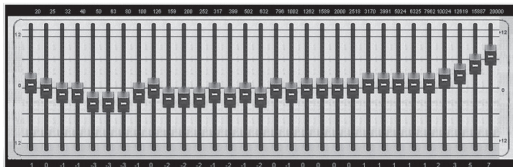


Abbildung 4.18: Grafischer Equalizer mit 31 Frequenzbändern im Abstand von jeweils einer Drittel Oktave

- Ein oder mehrere parallel geschaltete Bandfilter
- Einzelne Parameter jedes Filters können einzeln manipuliert werden
 - Mittenfrequenz
 - Güte (Steilheit der Flanken)
 - Faktor für Verstärkung/Dämpfung
- Benötigen mehr Erfahrung als einfache grafische Equalizer
- Sind in der Lage, komplexere Probleme zu lösen
 - Resonanzfrequenz eines Raumes

Dynamikkompressions

- Dynamik läßt sich als Verhältnis von größten und kleinsten Amplituden innerhalb eines Zeitfensters quantifizieren
- Je nach Länge des Fensters unterscheidet man Mikrodynamik (Sekundenbruchteile) und Makrodynamik (Sekunden und Minuten)
- Mikrodynamik macht z.B. den Charakter eines Instrumentes aus, Makrodynamik den Lautstärkeaufbau eines Musikstücks

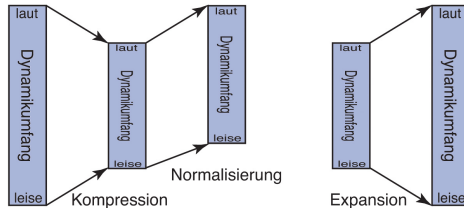
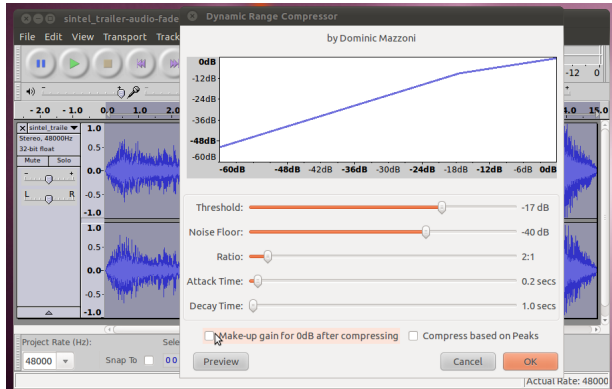


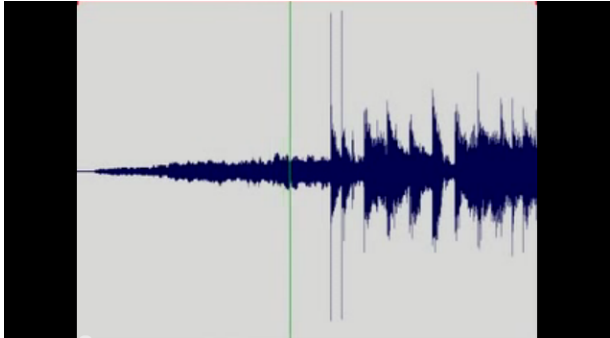
Abbildung 4.19: Kompression und Expansion der Dynamik eines Audiosignals

Audacity Compressor



Filtering and Effects

Video 6.1: Loudness Wars



👉 Matt Mayfield Music: Loudness Wars (1:53)

- Veränderung von Frequenz/Phase
 - Resampling: andere samplingrate (bewirkt Änderung der Signalfrequenz)
 - Time stretching: Länge des Signals verändern, ohne die Tonhöhe zu beeinflussen – kurze Abschnitte des Signals werden wiederholt (in Grenzen machbar)
 - Auch umgekehrt, z.B. um vorgeschriebene Hinweise in der Werbung zu verkürzen
 - Phasing/Flanging/Chorus: Analog zu den 3 Saiten eines Klaviers
- Echo und Hall
 - Echo: Original und zeitversetzte, abgeschwächte Reflexion
 - Hall: (mehrfach Reflexion), nachbildbar über Impulsantwort
- Restauration
 - Rauschen (Noise, Hiss), Klicken (Clicks), Knistern (Crackle) - Denoising über Fingerprint, Declicking über Interpolation des anderen Kanals

- Harter Schnitt: möglichst im Nulldurchgang, in Pausen

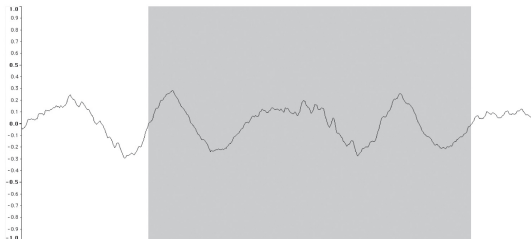


Abbildung 4.20: Markierung eines Signals zwischen zwei Nulldurchgängen

- Weicher Schnitt
 - Kreuz- oder Sturzblende

Audio

Jörg Cassens

Institut für Mathematik und Angewandte Informatik

Medieninformatik
WS 2017/2018



medieninformatik

IMAI – Institut für
Mathematik und
Angewandte Informatik

Alle Abbildungen, wenn nicht anders angegeben, aus Malaka et al. [2009].

Rainer Malaka, Andreas Butz, and Heinrich Hussmann.
Medieninformatik – Eine Einführung. Pearson Studium,
Munich, 2009.

John Watkinson. *The MPEG Handbook: MPEG-1, MPEG-2, MPEG-4*.
Technology & Engineering. Taylor & Francis, 2004.