

5. Ton und Klang

- 5.1 Ton: Physikalische und physiologische Aspekte
- 5.2 Kompression von Audio-Signalen: MPEG-Audio
- 5.3 Audio-Datenformate: Übersicht
- 5.4 Klangerzeugung und MIDI



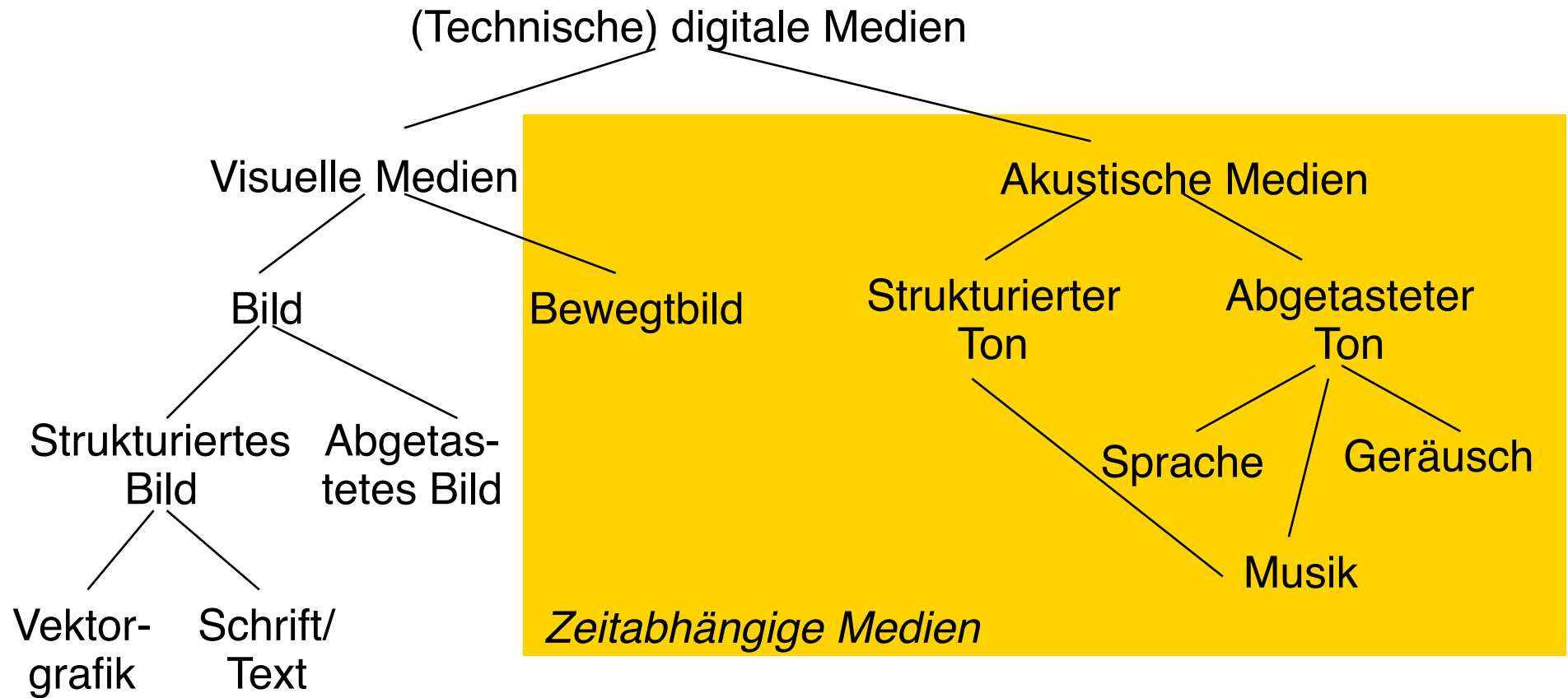
Literatur:

Medieninformatik-Buch:
Kapitel 4



Hannes Raffaseder: Audiodesign, Fachbuchverlag Leipzig 2002

Übersicht Medientypen



Weitere Klassifikationskriterien: komprimiert/unkomprimiert, ...

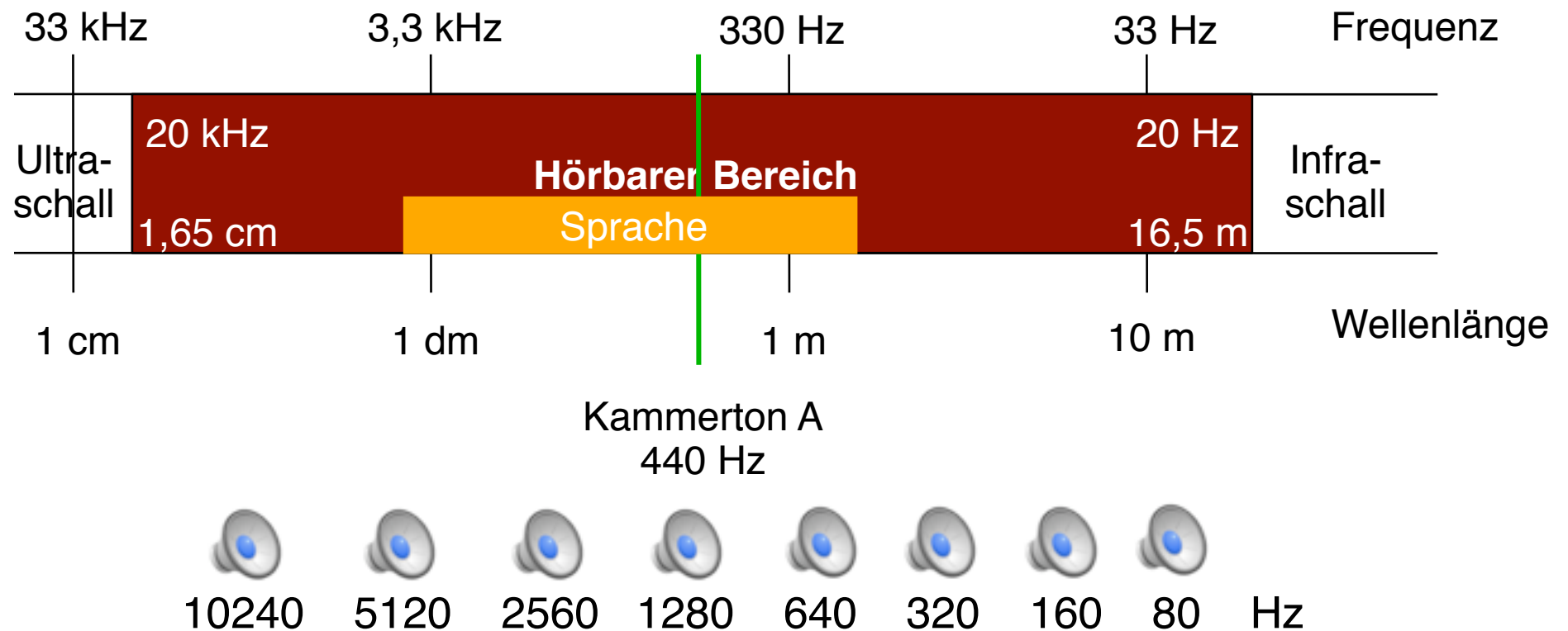
Angelehnt an Eidenberger & Divotkey

Übermacht der visuellen Eindrücke?

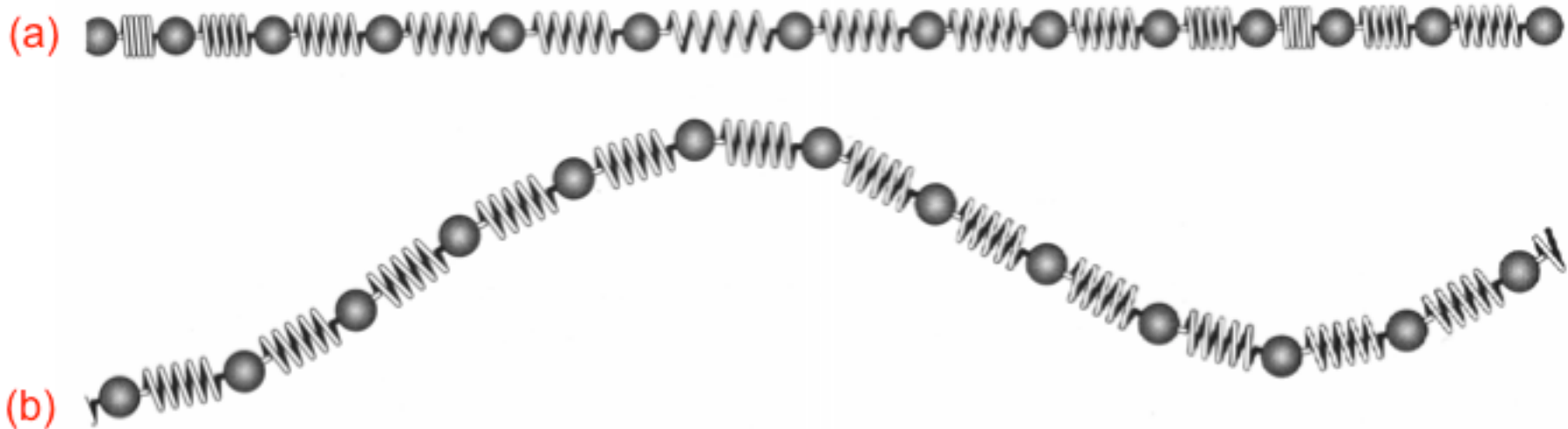
- Tendenz: Visuelle Gestaltung dominiert
 - CD-Cover, Bühnenshow, Interessenlage von Medieninformatik-Studierenden
 - Gibt das Auge den Ton an?
- Warum Gestaltung *akustischer* Medien?
 - Ohr nicht verschließbar, unbewusste Wahrnehmung als Normalfall
 - » Akustische „Szenerie“ bestimmt Grundstimmung, siehe Filmmusik
 - » „Ohrwürmer“
 - Direkte Kopplung der Schallverarbeitung mit Emotionen
 - » Feine „Untertöne“ in der Sprache
 - Verbindender, kollektivierender Effekt
 - » Blick auf Schallereignis, Jugendkultur, Marschmusik
 - Objektivierende Wirkung
 - » Akustische Ereignisse wesentlich schwerer zu verfälschen als optische
 - Aufmerksamkeitseffekt
 - » Akustische Alarme
- Audio-Gestaltung ist wesentlich in Kunst und Ingenieurwesen!

Ton (*sound*)

- Schallwellen, d.h. schnelle Änderungen des Luftdrucks
 - Longitudinalwellen, keine Schwingungsebenen
 - Ausbreitungsgeschwindigkeit ca. 331 m/s (bei 0°C), ca. 343 m/s (bei 20°C)
- Vom Menschen hörbares Spektrum ca. 20 Hz bis 20 kHz



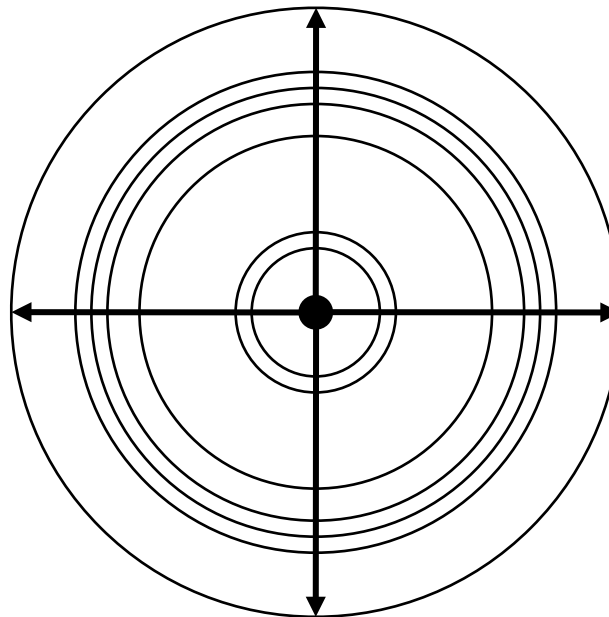
Longitudinal- und Transversalwellen



- Longitudinalwelle (a):
 - Oszillatoren schwingen in der Ausbreitungsrichtung
- Transversalwelle (b):
 - Oszillatoren schwingen quer zur Ausbreitungsrichtung
- Demo applet (http://www.cbu.edu/~jvarrian/applets/waves1/lontra_g.htm)

Schallausbreitung

- Schall breitet sich als *Kugelwelle* aus
 - Gleicher Schalldruck und gleiche Schallschnelle zu gegebenem Zeitpunkt auf konzentrischer Kugel um die Quelle
 - Bei größerer Entfernung: Äquivalent zu einer *ebenen Welle*



Wellenlänge

- Die Ausbreitung von Wellen (in diesem Fall Schallwellen) ist abhängig vom umgebenden Medium. Für jedes Medium gibt es eine spezifische Ausbreitungsgeschwindigkeit.
- Die Länge λ einer Welle (d.h. der Abstand bis zur nächsten Wiederholung) bestimmt sich aus der Periodenlänge T und der Ausbreitungsgeschwindigkeit c .

$$\lambda = c \cdot T = \frac{c}{f}$$

Reflexion

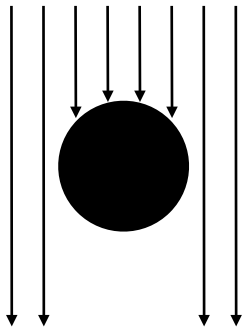
- Für Schallwellen gelten die gleichen Gesetze wie für Lichtwellen
- Aber: Effekte abhängig von Wellenlänge
 - Wellenlänge bei Schallausbreitung in der Luft bei 20°C zwischen ca. 20 m (tiefe Frequenzen) und 2 cm (hohe Frequenzen)
 - Lichtwellen erheblich kürzer (Mikrometer-Bereich)
- *Reflexion*:
 - Reflexionsgesetz (Einfallswinkel gleich Ausfallswinkel) gilt nur, wenn Grenzfläche groß im Vergleich zur Wellenlänge
 - » d.h. in kleineren Räumen keine Reflexion tiefer Frequenzen
 - Rauigkeit der Oberfläche führt zu *diffuser* (zerstreuender) Reflexion, wenn Unebenheiten in der Größenordnung der Wellenlänge
 - » d.h. auch bei zentimetergroßen Unebenheiten wirkt Wand auf tieffrequenten Schall als "glatt"

Absorption, Brechung

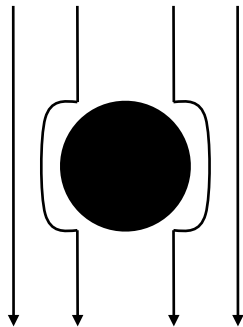
- *Absorption:*
 - Ein Teil der Schallenergie wird nicht reflektiert
 - » Absorptionsgrad abhängig vom Material und der Schallfrequenz
 - *Dissipation:* Umwandlung in Wärmeenergie
 - *Transmission:* Weiterleitung im absorbierenden Medium
 - Aufteilung der absorbierten Energie in Dissipation und Transmission
 - » frequenzabhängig: niedrige Frequenzen höhere Transmission
(Körperschall)
- *Brechung:*
 - Auch Schall wird gebrochen
 - Z.B. bei Luftschichten verschiedener Temperatur

Beugung

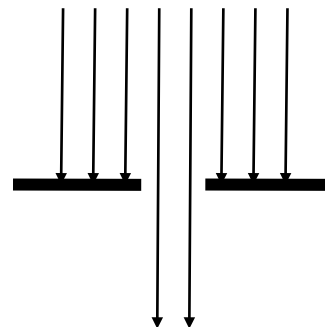
- Lichtwellen erzeugen immer einen Schatten hinter einem undurchsichtigen Gegenstand
 - Weil Gegenstand groß im Vergleich zur Wellenlänge
- Schall-"Schatten" entsteht erst, wenn ein Gegenstand groß im Vergleich zur Wellenlänge ist
- Bei Gegenständen in Größenordnung der Wellenlänge tritt *Beugung* auf
 - Wellen fließen "um das Hindernis herum"
 - Wellen werden von einem Spalt zerstreut



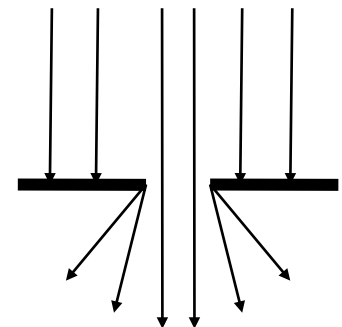
hochfrequenter Schall
an einer Säule



tieffrequenter Schall
an einer Säule

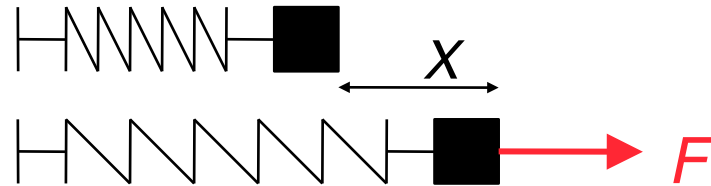


hochfrequenter Schall
an einem Spalt



tieffrequenter Schall
an einem Spalt

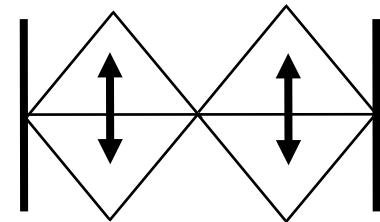
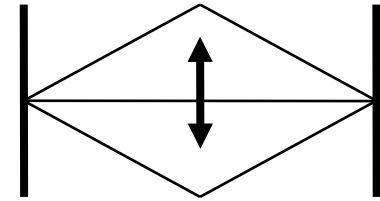
Schallentstehung: Schwingung



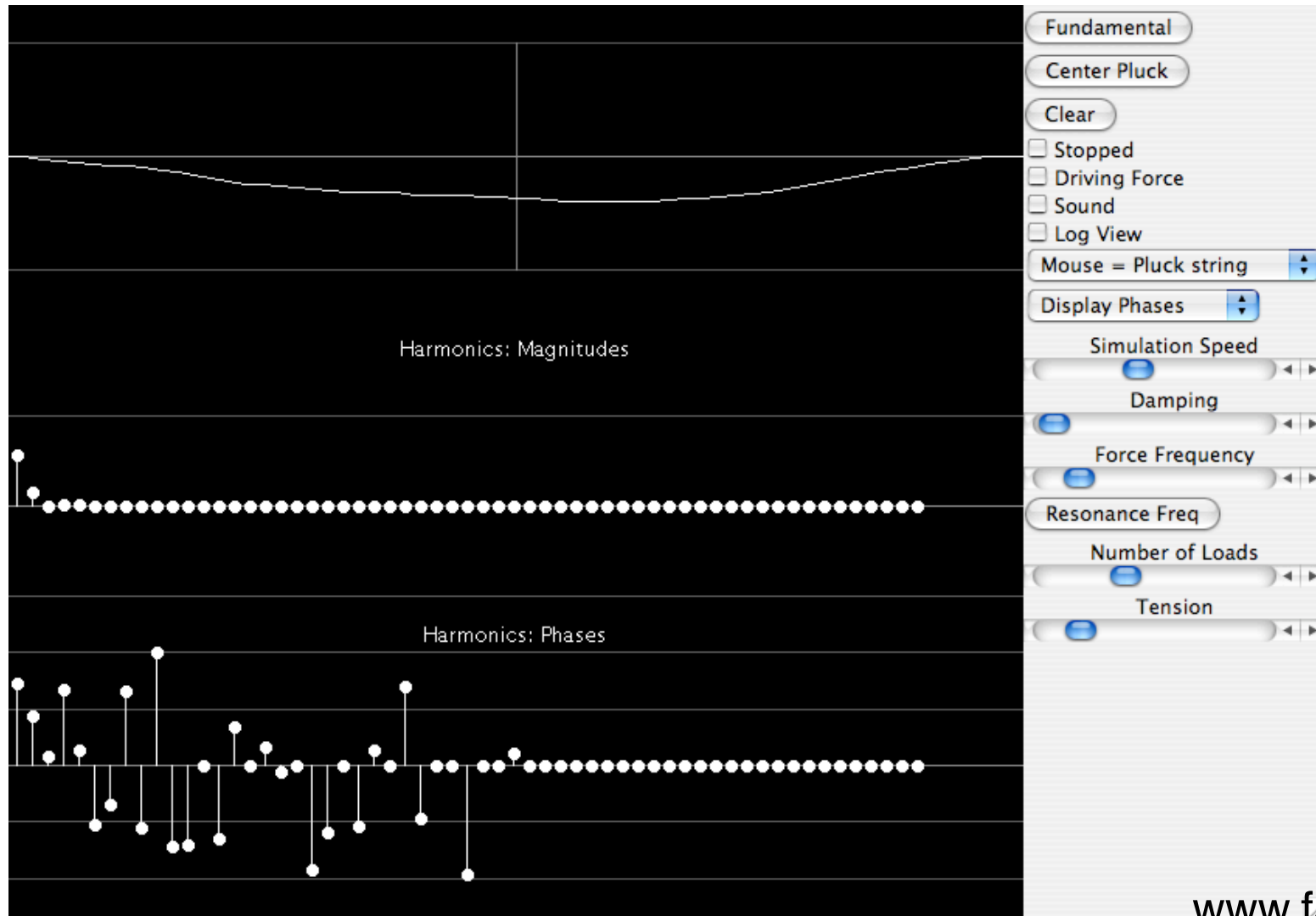
- *Anregungskraft*: Wirkt von außen auf Körper ein
- *Elastizität*: Beschreibt Verformbarkeit des Körpers
 - Beispiel oben: Feder
 - Bewirkt Tendenz eines verformten bzw. ausgelenkten Körpers, in die Grundstellung zurückzukehren
- *Massenträgheit*:
 - Masse kehrt nicht sofort in Ruhestellung zurück, sondern bewegt sich in Gegenrichtung der ursprünglichen Auslenkung weiter
 - Elastizität bewirkt entgegengesetzte Kraft
 - Gesamtergebnis: *Schwingung*
 - Frequenz umso höher, je kleiner die Masse und je geringer die Elastizität

Schwingungssysteme

- Schwingungssystem besteht aus:
 - Oszillator (z.B. Saite, Membran)
 - Anregung (z.B. zupfen, anstreichen, anschlagen)
 - Resonanzkörper (akustischer Verstärker, bestimmt Klangfarbe)
- Modell eines Schwingungssystems:
 - Gekoppelte einfache Schwingungssysteme (z.B. gekoppelte Federn)
 - » Konzeptionell auch unendlich viele gekoppelte Systeme (z.B. Modell einer Saite)
 - Randbedingungen
 - » z.B. eingespannte Saite: Auslenkung an den Enden immer Null
 - Begrenzte Anzahl von Wellenformen mit jeweils spezifischer Eigenfrequenz (“stehende Wellen”)
 - » Beidseitig eingespannte Saite der Länge L : Sinuswellen
 - » größtmögliche Wellenlänge $2 \cdot L$ (Grundfrequenz)
 - » ganzzahlige Vielfache der Grundfrequenz



Loaded string demo



www.falstad.com

Effektivwert und Pegel eines Signals

- Lautstärkenvergleich:
 - Wie bestimmen wir die „mittlere Amplitude“?
 - Arithmetischer Mittelwert falsch wegen negativer Werte
- *Effektivwert*:
 - Quadriert Signalstärkenwerte (und zieht am Ende die Wurzel)
 - Bildet Mittelwert über Zeitintervall

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

- *Pegel*:
 - Verhältnis zweier Größen
 - Riesige Wertunterschiede in der Schalldruck-Wahrnehmung des Menschen:
 - » Faktor 32×10^{12} (32 Trillionen) = 2^{45}
 - > *logarithmische Pegel*

Dezibel, Absolute Pegel





- Exponentieller Zusammenhang:
 - Verdoppelung eines physikalischen Basismaßes wird als eine Stufe (plus 1) aufgefasst (sog. logarithmisches Gesetz)
 - Beispiele für das logarithmische Gesetz:
 - » Empfindlichkeit des menschlichen Gehörs für Schalldruck
 - » Frequenzwahrnehmung in Oktaven
 - » Helligkeitwahrnehmung in Grauwerten
 - » Verlust in Telefonkabeln (als Funktion der Länge)
- Bel (B, benannt nach Graham Bell):
 - Logarithmisches Maß zu einem Referenzwert (0 Bel)
 - Referenzwert traditionell $1 \text{ mW an } 600 \Omega = 0,78\text{V} * 1,3\text{mA}$
- DeziBel (dB, „de-be“, „dee-bee“): $1 \text{ dB} = 1/10 \text{ B}$
- *Absolute Pegel*: Dezibel-Angabe relativ zu verschiedenen Referenzen:
 - Verlustleistung (m, W)
 - Schalldruck (SPL) relativ zu $2 \times 10^{-5} \text{ Pa}$ (Hörschwelle)
 - Bewerteter Schalldruck (A)

Relative Pegel

220Hz

Relativer *Leistungspegel* in dB:

$$L_P = 10 \cdot \log\left(\frac{p_A}{p_E}\right)$$

- 0 dB 
- 6 dB 
- 12 dB 
- 18 dB 

Verdopplung:

$$L_{P'} = 10 \cdot \log\left(\frac{2 \cdot p_A}{p_E}\right) = 10 \cdot \log(2) + L_P = 3 + L_P$$

Log₁₀(2)=0.3

Bezogen auf *Amplitudenpegel* (effektive Amplitudenwerte):

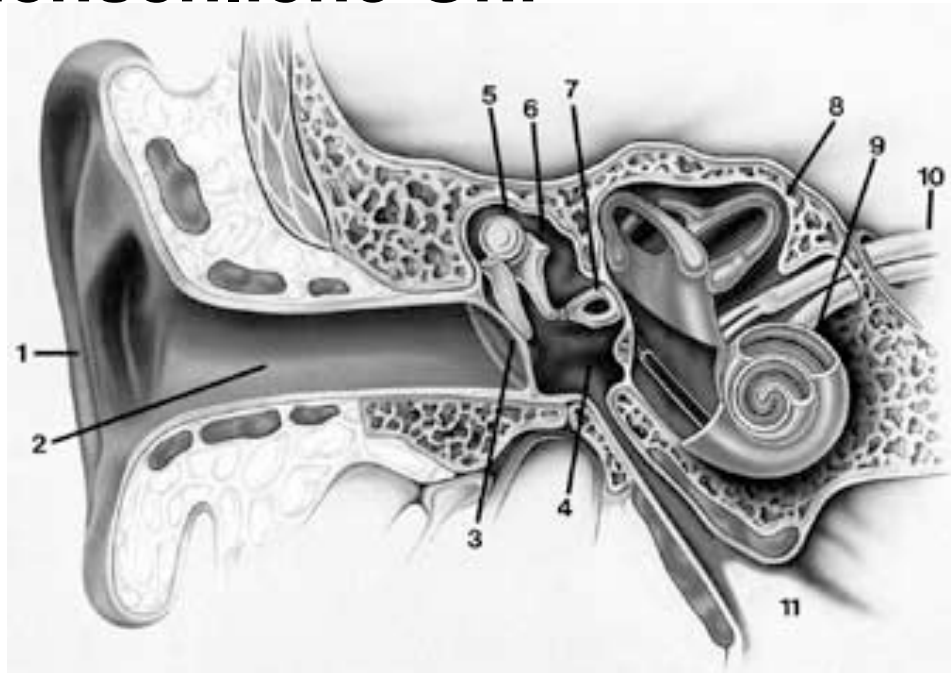
$$L_P = 10 \cdot \log\left(\frac{V_A^2}{V_E^2}\right) = 20 \cdot \log\left(\frac{V_A}{V_E}\right)$$

„Plus 6 dB
ist doppelte
Lautstärke“

Verdopplung:

$$L_{P'} = 20 \cdot \log\left(\frac{2 \cdot V_A}{V_E}\right) = 20 \cdot \log(2) + L_P = 6 + L_P$$

Das menschliche Ohr

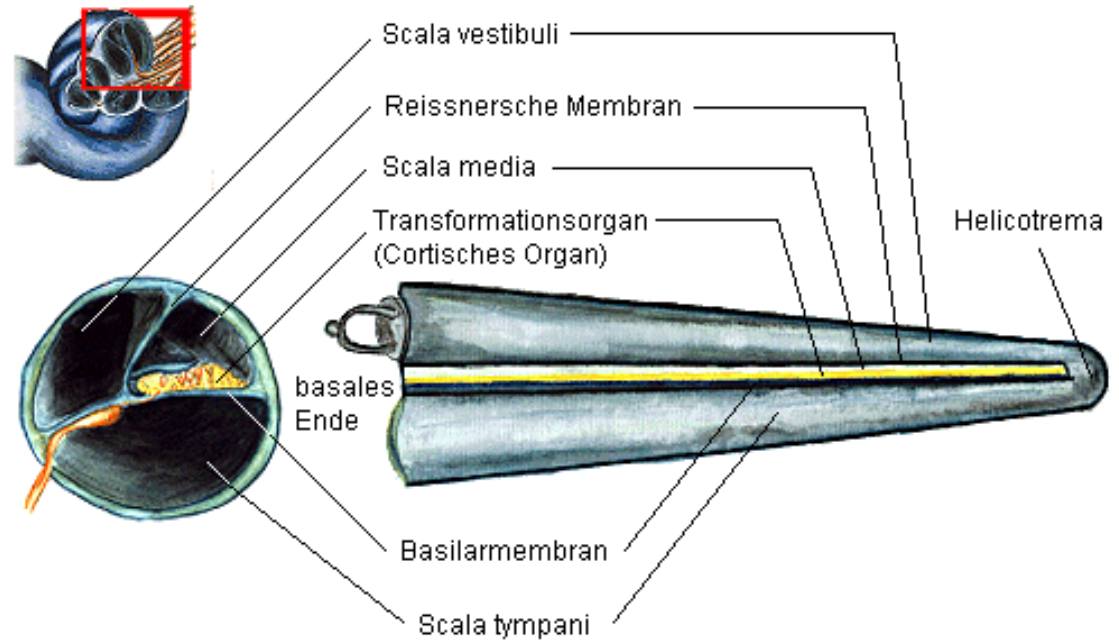


- 1 Ohrmuschel (*pinna*)
- 2 Gehörgang
- 3 Trommelfell (*ear drum*)
- 4 Mittelohr
- 5 Hammer (*malleus*)
- 6 Amboss (*incus*)
- 7 Steigbügel (*stapes*)
- 8 Gleichgewichtsorgan
- 9 Schnecke (*cochlea*)
- 10 Hörnerven

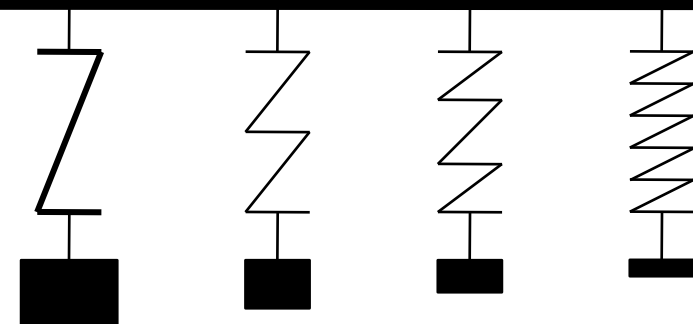
- Aussenohr:
 - Ohrmuschel
- Mittelohr:
 - Trommelfell, Hammer, Amboss, Steigbügel: Verstärkung der Kraftwirkung
- Innenohr:
 - Schnecke (*cochlea*)
 - Aufgerollte Röhren (Gänge), gefüllt mit Lymphflüssigkeit

Schnecke

- Aufgerollt: Im *abgerollten* Zustand ca. 30 mm lang
- Enthält 3 Röhren und die *Basilarmembran*
- Auf der Basilarmembran: *Corti'sches Organ*
 - 20000 Haarzellen
 - überträgt Bewegungen der Lymphflüssigkeit auf Haarzellen und damit auf Nervenreize
- Ort der Reizung im Corti'schen Organ abhängig von der Frequenz
 - konischer Aufbau und veränderte Elastizität der Membran zur Spitze hin



Vereinfachtes physikalisches Modell:

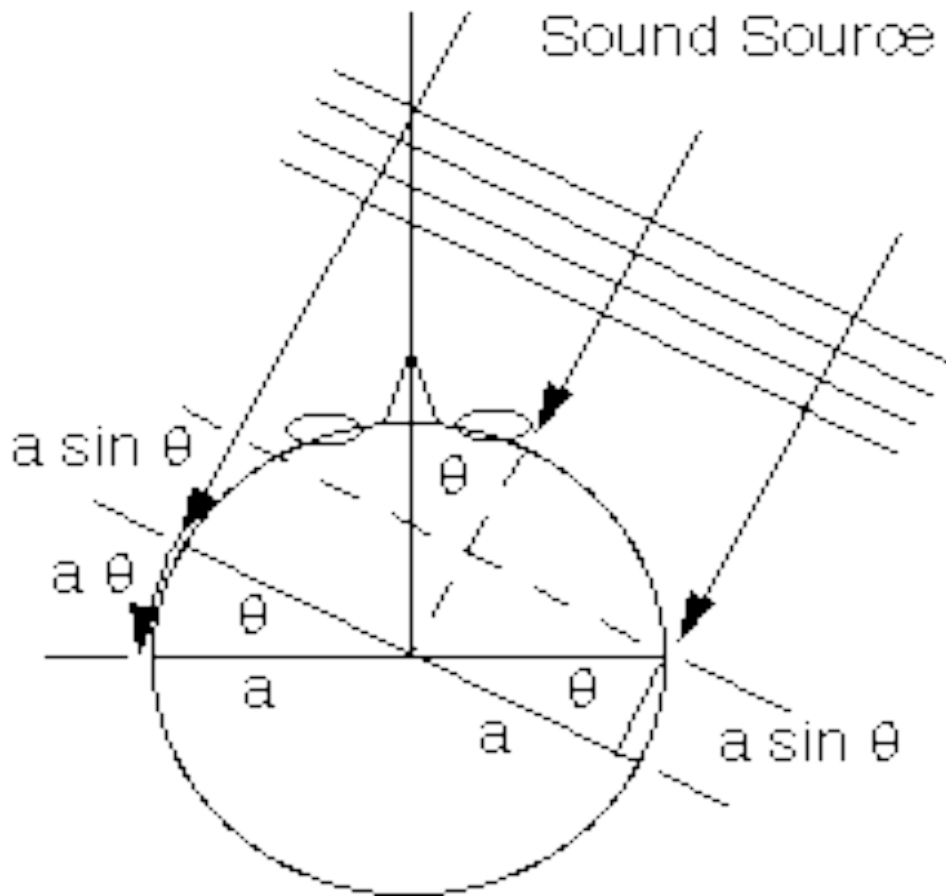


Räumliches Hören

- Stereo-Effekt
 - Unterschiedliche Lautstärke des Schallereignisses an beiden Ohren
 - Zeitliche Verzögerung des selben Schallereignisses in seiner Wahrnehmung durch beide Ohren
 - Verzögerungsmessung liefert Information über Entfernung der Quelle
- Kann ein einseitig tauber Mensch räumlich hören?
 - Eingeschränkt: ja!
 - Reflexion und Beugung an Umwelt und Ohrmuscheln liefern umfangreiche Information
- Frequenzabhängigkeit der Ortung:
 - Niedrige Frequenzen generell schlechter zu orten
 - Konsequenz physikalischer Tatsachen (Wellenlänge:Hindernis)
 - Siehe z.B. „Subwoofer“-Technologie

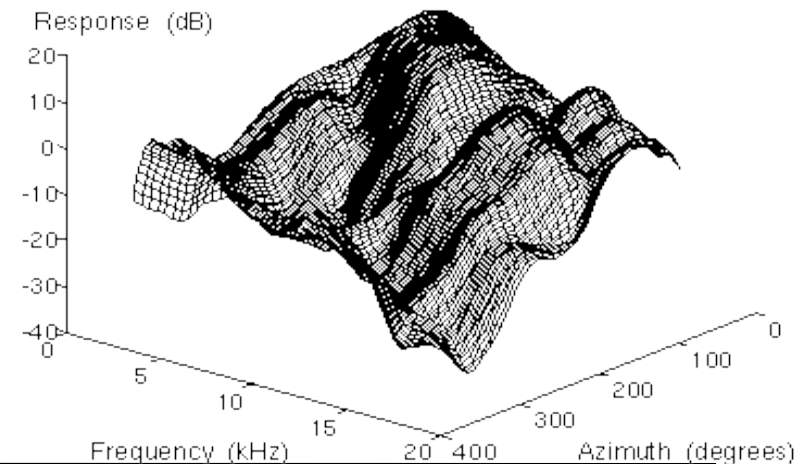
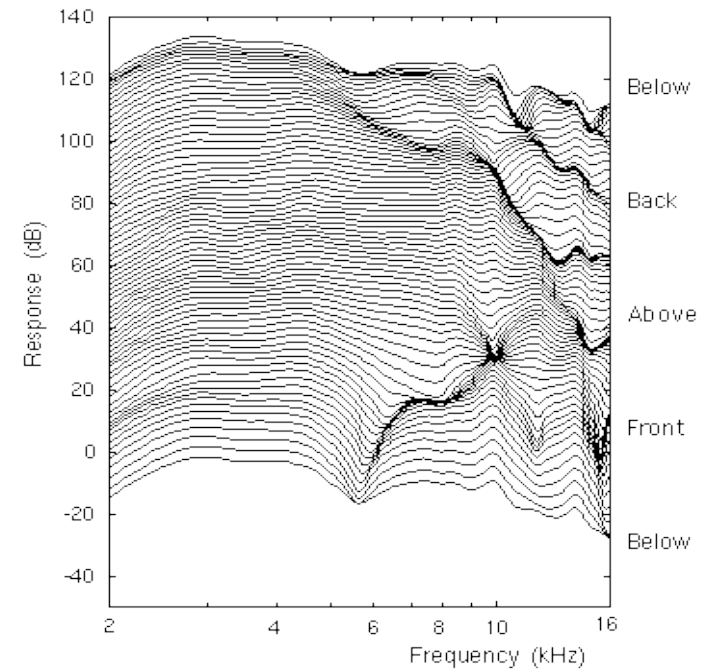
Hinweise (cues) für räumliches Hören

- Interaural intensity difference (IID)
- Interaural time difference (ITD)
- Head related transfer functions (HRTF)



Head Related Transfer Function

- Für alle Positionen rund um den Kopf Impulsantwort von der Position zu beiden Ohren messen → HRIR
- Fourier-Transformation davon ist die HRTF
- Enthält alle physikalischen cues für die Lokalisierung
- HRTF ist von Mensch zu Mensch verschieden
- Sobald HRTF für beide Ohren bekannt ist, kann mit einem Kopfhörer Raumklang (also auch vorne/hinten und oben/unten) erzeugt werden



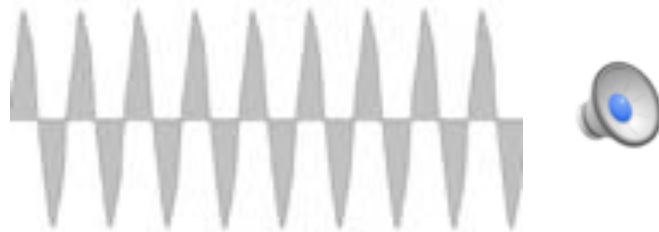
Empfindungen: Klang und Geräusch

- Töne sind vom Menschen wahrnehmbare kleine Luftdruckänderungen
 - Warum empfinden wir manche Tonereignisse als "Musik", als Klang?
- Primärempfindungen der Tonwahrnehmung:
 - Tonhöhe (Bsp. verschiedene Klaviertasten)
 - Lautstärke (Bsp. Trommelanschlag)
 - Klangfarbe (Bsp. gleicher Ton auf verschiedenen Instrumenten)
- Klang:
 - alle drei Primärempfindungen wahrnehmbar
- Geräusch:
 - nur Lautstärke wahrnehmbar



Periodizität

- Klänge sind, als Signalform betrachtet, *periodisch* (d.h. wiederholen Teilabschnitte)
- Geräusche sind schlechter strukturiert und meist aperiodisch.



Blockflöte



LKW



Violine



Fahrradkette

Frequenzspektrum, Oberschwingungen

- *Frequenz* (Tonhöhe):
 - Maß für die Häufigkeit, mit der sich positive und negative Spannungen abwechseln, Maß 1 Hertz = 1 Schwingung/s
- Audiosignal:
 - besteht aus Vielzahl von überlagerten Frequenzen (Frequenzspektrum)
 - *Bandbreite*: Differenz zwischen höchster und niedrigster Frequenz
 - Beispiel: Ton eines Musikinstrumentes
- *Grundton*: Wahrgenommene Tonhöhe
 - Größter gemeinsamer Teiler aller am Signal beteiligten Frequenzen
 - Oft: Tiefste enthaltene Teilfrequenz
- *Obertöne*:
 - *reine Obertöne*: ganzzahlige Vielfache der Grundfrequenz
 - zusätzlich: geräuschhafte Tonanteile (z.B. Zupfgeräusch)
 - Obertonspektrum ist für charakteristischen Instrumentklang bestimmend

Demo applet

Harmonische Schwingungen

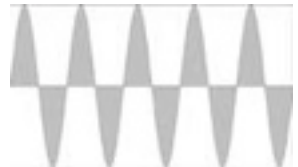
$$s(t) = \sum_{k=1}^{\infty} a_k \sin(k \cdot 2\pi \cdot f \cdot t + \theta_n)$$

- *Harmonische Schwingung (harmonischer Ton):*
 - Beschreibbar durch Sinus- und Cosinusfunktionen
- *Harmonisch komplexe Schwingung (Klang):*
 - Zusammengesetzt aus harmonischen Teilschwingungen
 - Grundton und ganzzahlige Vielfache
- *Klänge ohne Grundton:*
 - Zusammengesetzt aus harmonischen Teilschwingungen
 - Keine ganzzahligen Vielfachen einer Grundfrequenz
 - Z.B. Pauken, Gongs

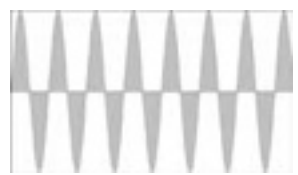
Überlagerung harmonischer Schwingungen



Sinus 110 Hz
(Grundton)



Sinus 220 Hz
(Oberton)



Sinus 330 Hz
(Oberton)



110 Hz + 220 Hz



220 Hz + 330 Hz



110 Hz + 220 Hz + 330 Hz



Demo applet

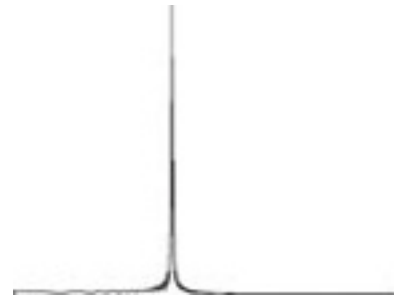
Interferenz und Schwebung

- *Interferenz*: Überlagerung von Schallwellen exakt gleicher Frequenz
 - konstruktive Interferenz (*in phase*):
 - » Übereinstimmung der Phasenlage
 - » Addition der Amplituden
 - destruktive Interferenz (*out of phase*):
 - » Gegenphasige Lage (180° verschoben)
 - » Subtraktion der Amplituden – Auslöschung
- *Schwebung*: Überlagerung von Wellen annähernd gleicher Frequenz
 - konstruktive und destruktive Interferenz wechseln sich ab
 - Amplitudenverlauf beschreibt neues Signal mit Frequenz = Differenz der überlagerten Frequenzen
- Diskussion: Was passiert, wenn ich bei meiner Stereoanlage einen Lautsprecher richtig, den andern mit +/- vertauscht anschlieÙe?

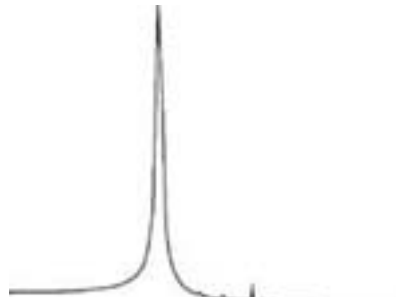
Frequenzspektren

- Frequenzspektrum von Klängen
 - Anzahl diskreter Spektrallinien (Grund- und Obertöne)
- Frequenzspektrum von Geräuschen
 - kontinuierliches Spektrum diverser Frequenzen

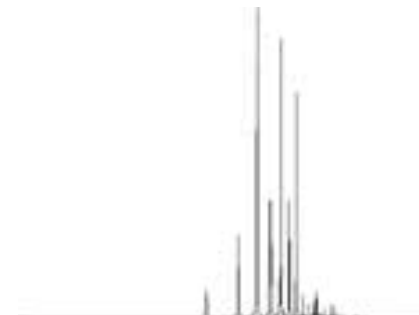
Klänge:



Sinus 110 Hz



Blockflöte

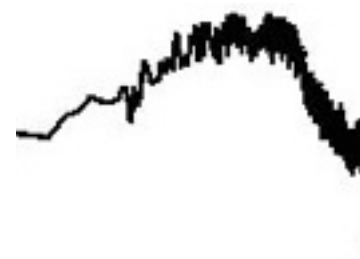


Violine

Geräusche:



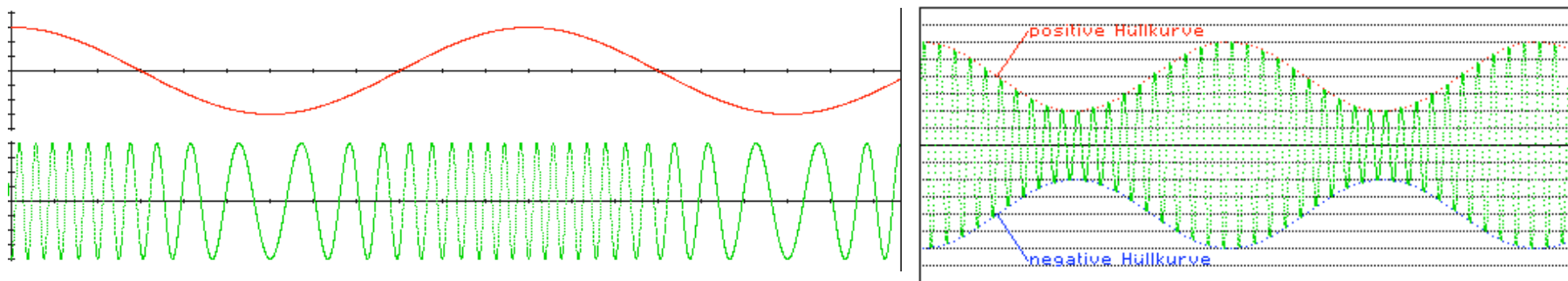
Bach



Fahrrad (Bremsen)

Modulation

- Modulation:
 - gezielte Überlagerung einer Grundfrequenz (Träger) mit einem Nutzsignal
 - Anwendung in der Übertragungstechnik (z.B. Rundfunk)
- Amplituden-Modulation (AM):
 - Hüllkurve der Trägerfrequenz (Amplitude) in Abhängigkeit vom Nutzsignal verändert
- Frequenz-Modulation (FM):
 - Abweichung von der Basisfrequenz in Abhängigkeit vom Nutzsignal
- Diskussion: Warum bedeutete beim analogen Rundfunk der Wechsel von AM auf FM eine Qualitätsverbesserung?



Zeitlicher Verlauf von Schallsignalen, Transienten

- Mikroskopischer Zeitbereich: $0,05 \text{ ms} < t \leq 50 \text{ ms}$
 - Hörbare Frequenzen (20 Hz bis 20 kHz)
- Zeitbereich der Übergangsklänge (*Transienten*): $50 \text{ ms} < t \leq 150 \text{ ms}$
 - Modulation hörbarer Frequenzen
 - So "schnell" (20 Hz bis 7 Hz), dass noch als Variation des Klangs wahrgenommen
- Makroskopischer Zeitbereich: $t > 150 \text{ ms}$
 - Verlauf eines Klangs (z.B. gespielte Note) über die Zeit (Hüllkurve)
 - Formaler Aufbau eines Musikstücks
 - Tempo, Metrum, Rhythmus
- Beispiele für Transienten:



Informationsgehalt akustischer Szenen

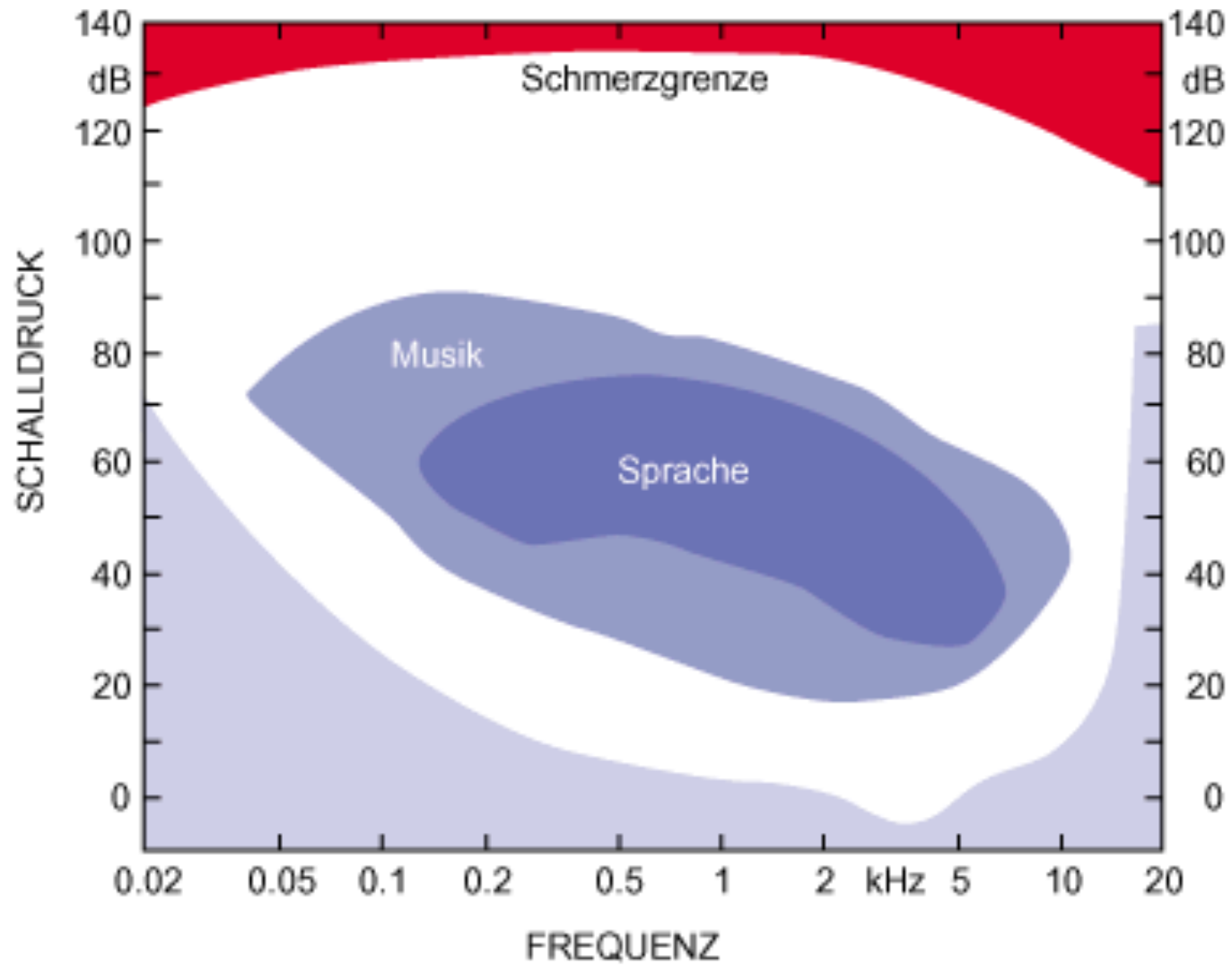
- Umgebung des Menschen ist durch Vielzahl verschiedenartiger Geräusche geprägt
- Menschlicher Gehörsinn (= Hörapparat + Nachverarbeitung)
 - Kann gezielt Aufmerksamkeit auf sehr spezifische Teilsignale richten:
Cocktailparty-Effekt
 - Nimmt Hintergrundinformationen neben selektierter Information weiter wahr (z.B. Alarmsignale, Durchsagen)
 - Kann räumlich weit verstreute Informationsquellen integrieren (wesentlich geringere Verdeckung als in der Optik)
 - Kann sehr genau verschiedene Schallereignisse differenzieren („Mischfarben“ vs. Wahrnehmung gleichzeitiger Töne)
- Digitale Tonverarbeitung muss sich mit der Informationsverarbeitung im Gehirn befassen, nicht nur mit dem physikalischen Hörorgan

Psychoakustik

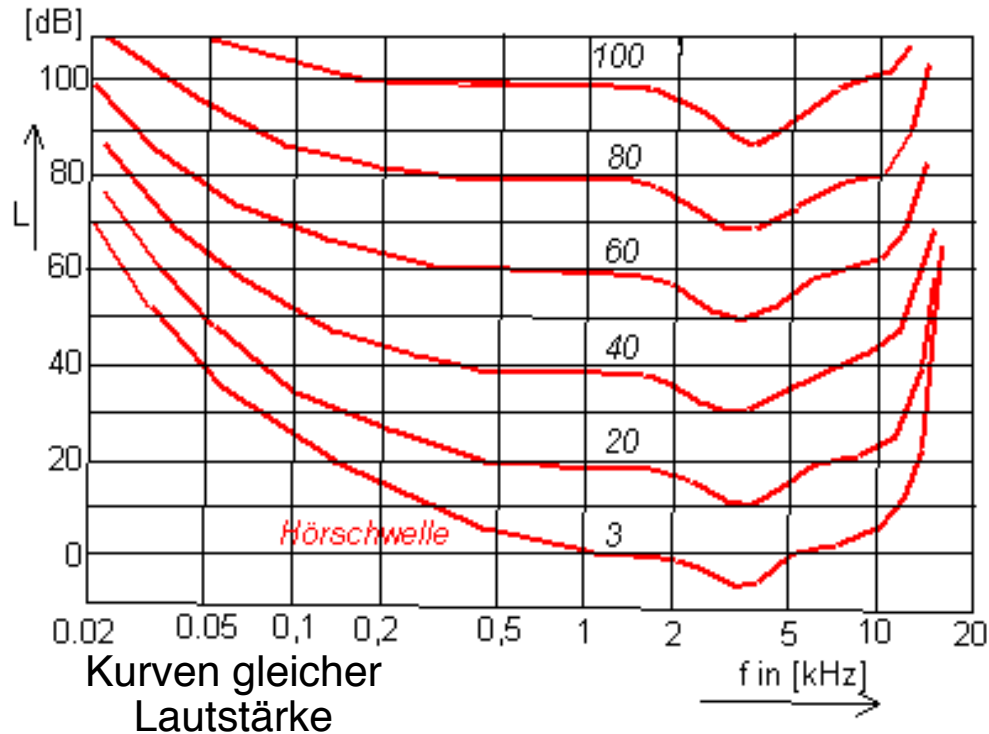
- Die *Psychoakustik* versucht kausale Zusammenhänge zwischen den physikalischen Größen eines Schallsignals und den dadurch ausgelösten Empfindungen zu erfassen.

Mensch wahrnehmbare Eigenschaft	Computer physikalische Kenngröße
Tonhöhe	Grundfrequenz
Lautstärke	Druckamplitude
Klangfarbe	Frequenzspektrum

Hörfläche



Lautstärke und Frequenz

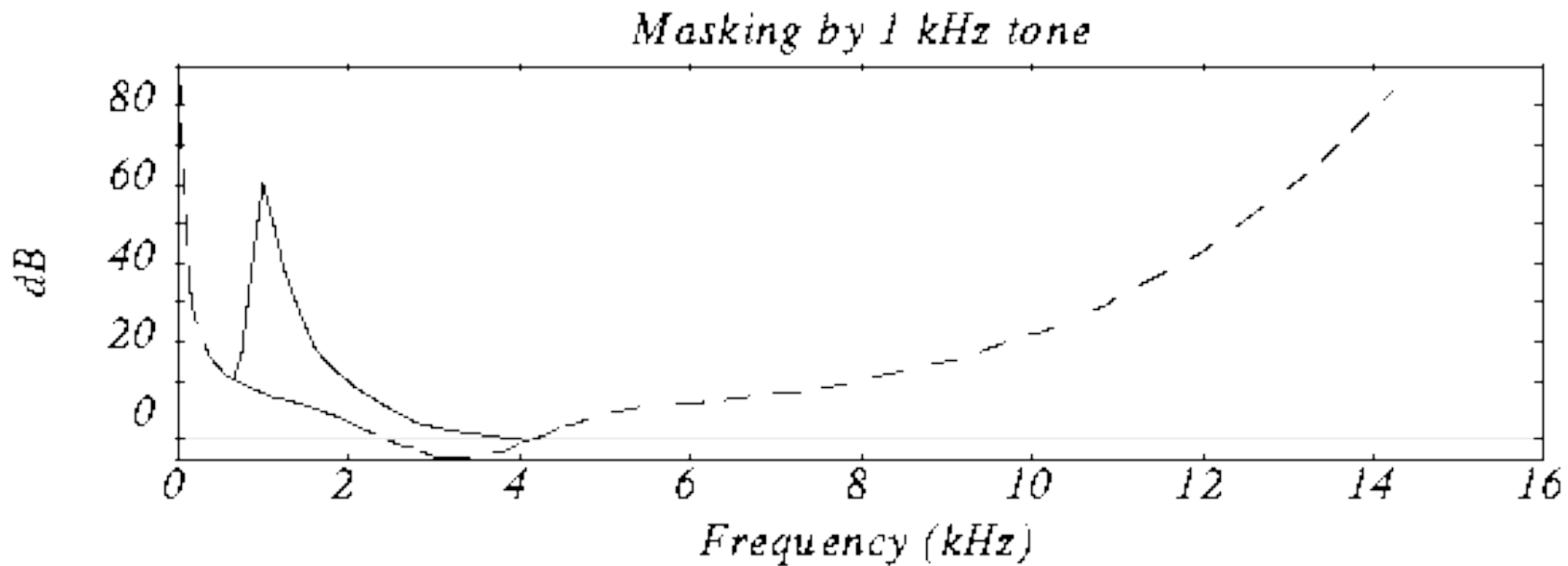


- Die Hörschwelle ist abhängig von der Frequenz:
 - Größte Empfindlichkeit bei ca. 2 – 5 kHz
- Gleich empfundene Lautstärke frequenzabhängig bei verschiedenem Schalldruck
 - *Subjektive* Lautstärke gemessen in *Phon*
 - x Phon: Lautstärke vergleichbar zu 1 kHz Ton bei Schalldruck von x dB (SPL)

- Korrektur der Frequenzabhängigkeit durch Gewichtungskurven (A, B, C)
 - meistverwendet A-Kurve (40 Phon), dB(A)
 - für höhere Lautstärken B- und C-Kurven
- „Loudness“: Bei geringer Lautstärke Anhebung Bässe u. Höhen

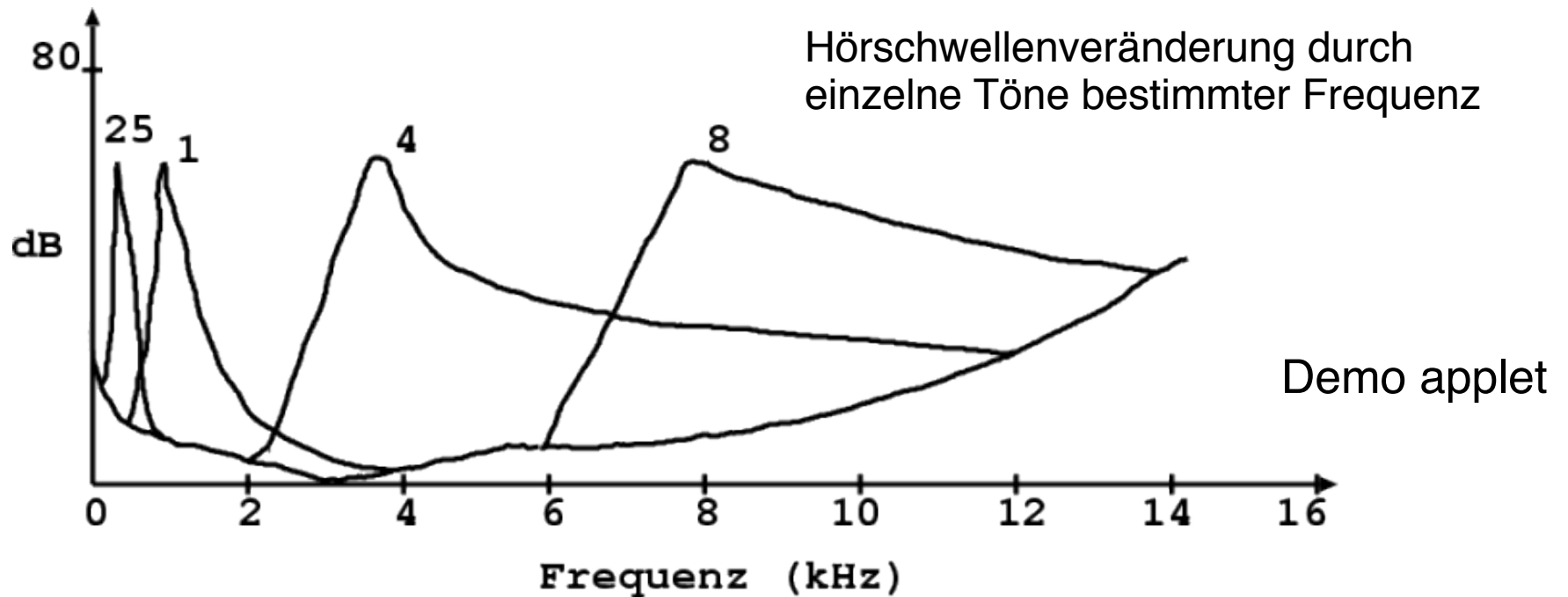
Frequenz-Maskierung (1)

- **Ein lauter Ton einer bestimmten Frequenz macht leisere Töne ähnlicher Frequenz unhörbar.**
- Experiment: Ein „Maskierungston“ (z.B. 1 kHz) wird mit festem Pegel (z.B. 60 dB) abgespielt. Ein zweiter „Testton“ (z.B. 1.1 kHz) wird mit einem anderen Pegel abgespielt, der so hoch ist, dass man die beiden Töne gerade unterscheiden kann.
- Für variierende Frequenzen des Testtons wird aufgezeichnet, ab welchem Pegel der Testton hörbar wird.



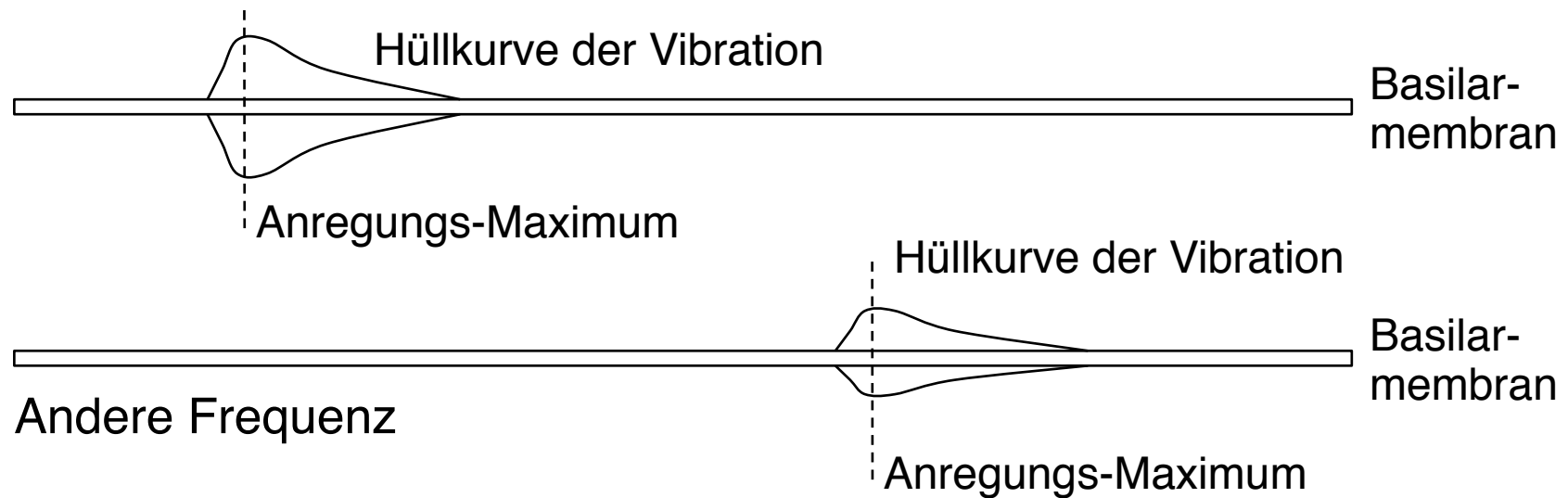
Frequenz-Maskierung (2)

- Frequenzwahrnehmung durch die Schnecke:
 - endliche Breite des betroffenen Bereichs
 - dadurch Überlappung benachbarter Frequenzbereiche



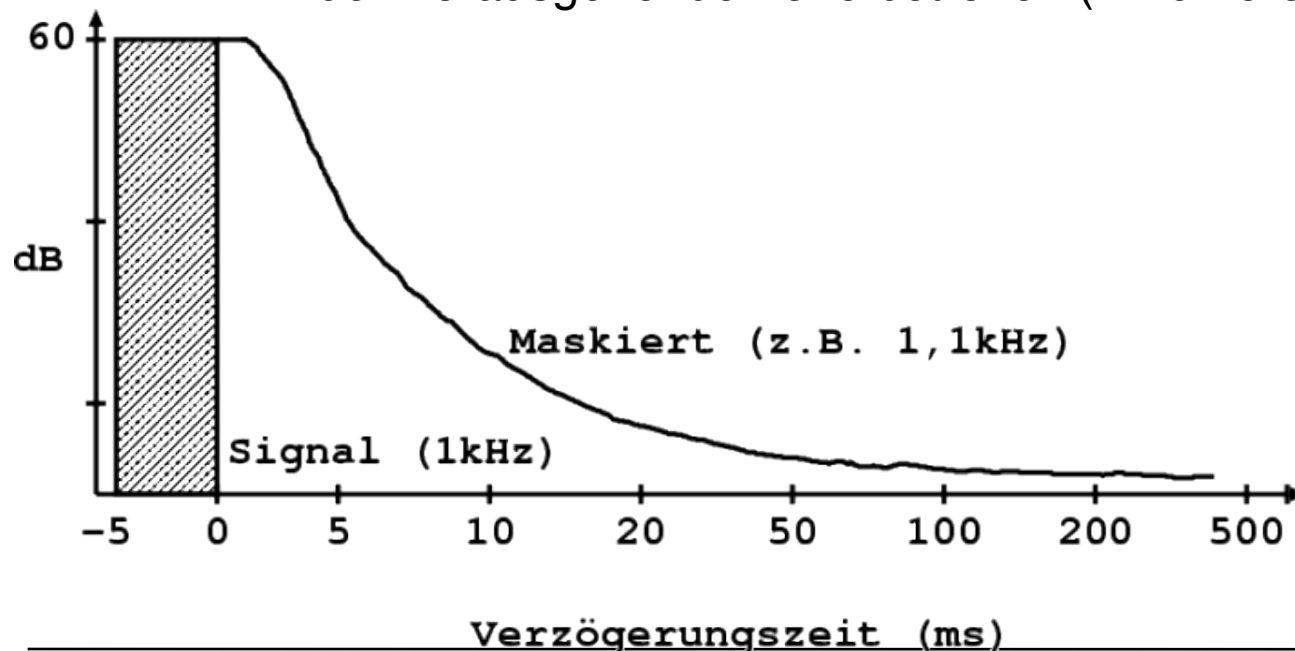
Maskierung und Basilarmembran





- Der Maskierungseffekt erklärt sich physikalisch durch die Anregung der Basilarmembran
 - Frequenz entspricht Ort der Anregung auf der Basilarmembran
 - Genaue Wahrnehmung des Maximums der Anregung (Auflösung ca. 1/12 Halbton, bestimmt durch Abstand der Haarzellen)
 - Anregungen in direkter Frequenz-Nähe sind bis zu einer bestimmten Amplitude nicht wahrnehmbar



Zeitliche Maskierung

- Die Hörwahrnehmung beruht auf dem Mittelwert eines Zeitintervalls von ca. 30 ms
 - Ein lauter Ton beeinflusst die Wahrnehmung einer frequenzähnlichen Tons auch, wenn der zweite Ton in direkter zeitlicher Nachbarschaft liegt
 - Vorwärtsmaskierung: Nachfolgende Töne kaum wahrnehmbar
 - Rückwärtsmaskierung:
 - » Auch vorausgehende Töne betroffen (in kürzerem Zeitabstand)



 Vorwärts, 100 ms
 Vorwärts, 10 ms
 Rückwärts, 100 ms
 Rückwärts, <10 ms

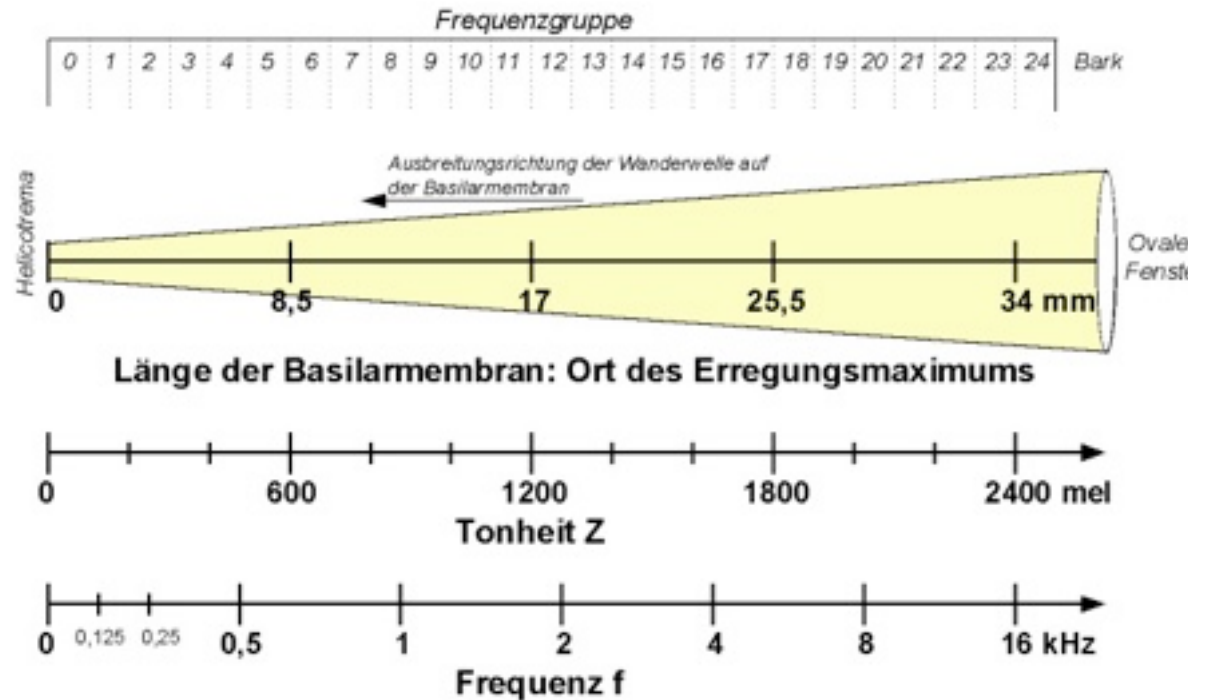
Demo:
<http://www.ece.uvic.ca/~upward/p/demos.htm>

Kritische Bänder

- Einteilung des Hörbereichs in *kritische Bänder*
 - Breite (d.h. Bandbreite im Frequenzspektrum) der Vibrations-Hüllkurve auf der Basilarmembran
 - Breite der Bänder vergrößert sich mit der mittleren Bandfrequenz
- Der Grad der Maskierung einer bestimmten Frequenz ist lediglich abhängig von der Signalintensität im kritischen Band dieser Frequenz.
- "Bark-Skala":
 - Einteilung des Frequenzspektrums entsprechend der Breite kritischer Bänder
 - Benannt nach dem Bremer/Dresdner Physiker Heinrich Barkhausen.

24 (oder 27) Kritische Bänder

0 – 100	630 – 770
100 – 200	770 – 920
200 – 300	920 – 1080
300 – 400	1080 – 1270
400 – 510	1270 – 1480
510 – 630	1480 – 1720
1720 – 2000	4400 – 5300
2000 – 2320	5300 – 6400
2320 – 2700	6400 – 7700
2700 – 3150	7700 – 9500
3150 – 3700	9500 – 12000
3700 – 4400	12000 – 15500



Bildquelle: Wikipedia
Tonheit: Maß für wahrgenommene
 Tonhöhe

5. Ton und Klang

- 5.1 Ton: Physikalische und physiologische Aspekte
- 5.2 Kompression von Audio-Signalen: MPEG-Audio
- 5.3 Audio-Datenformate: Übersicht
- 5.4 Klangerzeugung und MIDI



Weiterführende Literatur:

Arne Heyda, Marc Briede, Ulrich Schmidt: Datenformate im Medienbereich, Fachbuchverlag Leipzig 2003, Kapitel 5

John Watkinson: MPEG Handbook, 2nd ed., Butterworth-Heinemann 2004

Pulse Code Modulation (PCM)

- Klassische Digitalisierung:
 - Aufzeichnung des analogen Signalwertes zu festgelegten Zeitpunkten mit festgelegter Auflösung
- G.711 (für Telefonie):
 - 8 kHz Abtastfrequenz für 4 kHz breites Teilband (Sprache)
 - Auflösung 8 bit
 - 64 kbit/s Bandbreite = Breite eines ISDN „B-Kanals“
- Viele weitere Anwendungen
 - z.B. digitale Tonaufzeichnung auf Videoband (PCM-1630)

- Kompression von Audiodaten
 - Verlustfreie Kompression nur wenig wirksam
 - Generell relativ niedrige Kompressionsraten erreichbar

Verlustbehaftete Audio-Kompressionsverfahren

- Verlustbehaftete Audiokompression
 - Basiert auf psychoakustischem Modell der Tonwahrnehmung
 - Wichtigster Effekt:
Maskierte Bestandteile des Audio-Signals werden nicht codiert
 - Bekanntester Standard: MPEG Audio Layer III (MP3)
- MPEG = Moving Picture Expert Group
 - Standardisierungsgremium von ISO (International Standards Organization) und IEC (International Electrotechnical Commission)
 - Arbeitet seit 1988 an Video- und Audio-Kompression
 - » Untergruppe MPEG/Audio
 - MPEG-Audio-Standards werden z.B. verwendet bei
 - » DAB (Digital Audio Broadcast)
 - » DVB (Digital Video Broadcast) incl. terrestrischer Variante DVB-T
 - » DVD-Video

MPEG Audio: Geschichte

- EU-gefördertes "Eureka"-Projekt Nr. 147 (CCETT(F), IRT(D), Philips (NL))
 - MUSICAM (Masking pattern adapted universal sub-band integrated coding and multiplexing)
 - Ziel: DAB-Standard
- Parallelentwicklung (AT&T, Thomson, Fraunhofer, CNET):
 - ASPEC (Adaptive Spectral Perceptual Entropy Coding)
 - Ziel hochwertiges Audio über ISDN
- Juli 1990: Ausführliche Tests beim Schwedischen Rundfunk, anschließend Kombination der beiden Verfahren in die 3 MPEG-Layer.
 - Layer I: vereinfachtes MUSICAM, schwache Kompression, preisgünstig
 - Layer II: = MUSICAM, für DAB und Audio in DVB
 - Layer III: Kombination der Stärken von ASPEC und MUSICAM, hohe Kompression über Telekommunikationsverbindungen

MPEG: Video und Audio

	MPEG-1 Video Standard	MPEG-2 Video Standard
MPEG Audio Layer I		
MPEG Audio Layer II		
MPEG Audio Layer III	MP3	

Mehr Varianten

MP3 = MPEG Audio Layer III

MP3 ≠ MPEG-3 !!!

Audio-Codierung in MPEG

- MPEG-1 Audio:
 - PCM mit 32, 44.1 oder 48 kHz
 - max. Datenrate 448 kbit/s
- MPEG-2 Audio:
 - PCM mit 16, 22.05, 24, 32, 44.1 oder 48 kHz
 - max. 5 Kanäle
 - max. Datenrate 384 kbit/s
- Einteilung der Audio-Kompressionsverfahren in drei „Layer“ (I, II, III) verschiedener Kompressionsstärke
 - Unabhängig von Wahl des Standards MPEG-1 bzw. MPEG-2 !
 - „MP3“ = MPEG Layer III (Kompression ca. 11:1)
 - » MP3 patentrechtlich geschützt, Fraunhofer IIS Erlangen
- Inzwischen wesentliche Weiterentwicklungen:
 - z.B. AAC, MPEG-4 Audio (siehe später)
 - Ogg-Vorbis

Referenzmusik:

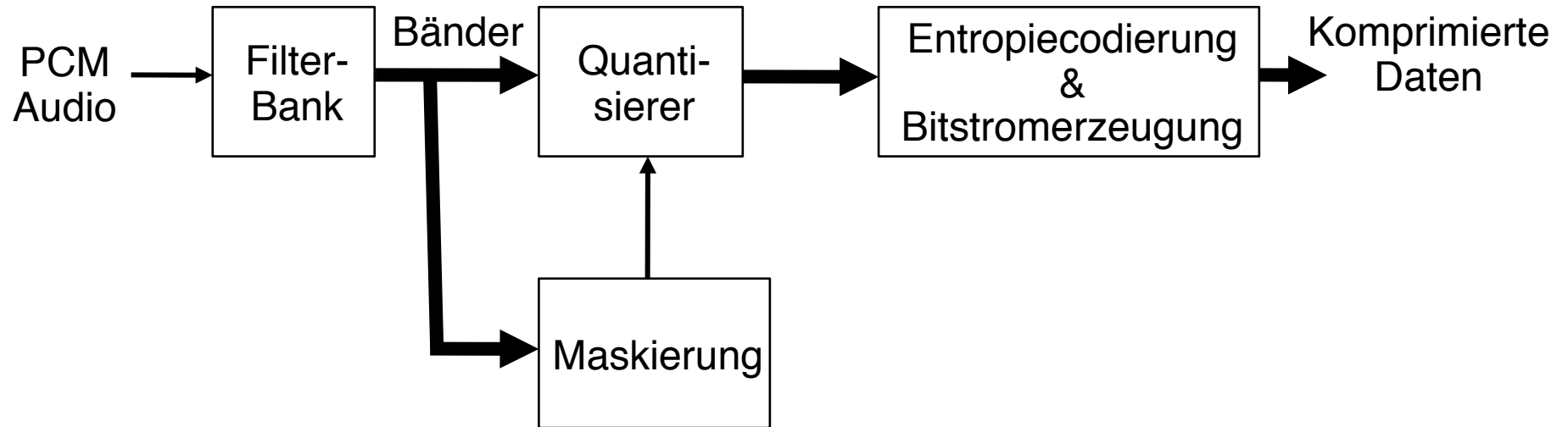
Tom's Diner (Suzanne Vega)

"I was ready to fine-tune my compression algorithm...somewhere down the corridor, a radio was playing 'Tom's Diner.' I was electrified. I knew it would be nearly impossible to compress this warm a cappella voice."

K.-H. Brandenburg Interview in Business 2.0



MPEG-Audio Encoder: Grundlegender Aufbau

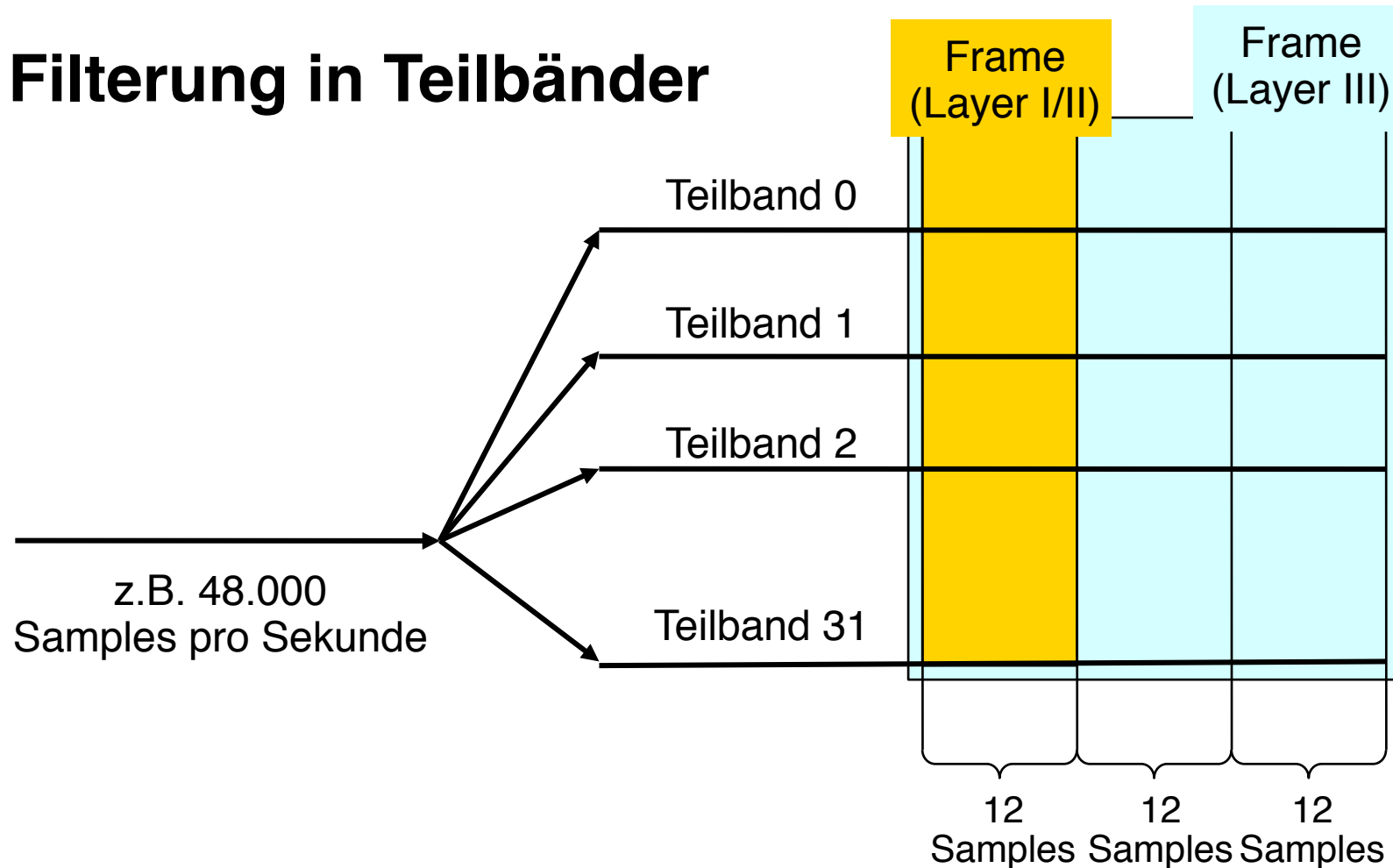


- Hinweis: Der MPEG-Standard definiert nicht den Aufbau eines Encoders, sondern nur die Decodierung!
- Signal wird in Frequenzbänder aufgeteilt
- Maskierung auf der Basis der Bänder mit einem psychoakustischen Modell

Subband-Kodierung

- Energie eines Tonsignals ist meist nicht gleichmäßig auf das Frequenzspektrum verteilt
- Idee:
 - Aufteilen des Signals in Teil-Frequenzbänder
 - Ermittlung des Signalpegels für jedes Teilband
 - Einzel-Codierung der Teilbänder mit jeweils angemessener Bitanzahl
 - » z.B. nicht belegtes Teilband: 0 Bit
 - Funktioniert optimal, wenn Teilbänder an kritische Bänder des Gehörs angepasst

Filterung in Teilbänder



- 12 Samples entsprechen bei 48 kHz ca. 8 ms
- Ein Block von Samples in einem Teilband wird manchmal *bin* genannt
- *Frame*: Gesamtheit der Samples in allen Teilbändern
 $12 \times 32 = 384$ Samples in Layer I/II, $3 \times 12 \times 32 = 1152$ Samples in Layer III