

SDV2: Datenaufnahme Audio

Schall wird durch bewegende Materie erzeugt (z.B. Gitarrensaiten)

↳ breitet sich in einem Medium (z.B. Luft aus)

↳ Schallwellen bringen Moleküle des Mediums in Schwingung

↳ in Richtung der Ausbreitungsrichtung

↳ nennt man auch Längswellen oder Longitudinalwellen

Luftschall sind Luftdruckänderungen

Schall:

- Geschwindigkeit abhängig von Dichte und Hitze des Mediums

- Ausbreitung: 1-2m kugelförmig

Verdoppelung des ~~Abstands~~ Abstands:

- Vervielfachung der Fläche
- Dezibel -6 dB

Geschlossener Raum: Nachhall durch Reflektionen

↳ 2. Reflektionen von Schallwellen im geschlossenen Raum

Unterschied Echo: Starke Verzögerung $\Rightarrow$ separates Hörereignis

Mikrofone wandeln Schalldruckänderungen (dB) in entsprechende elektrische Spannungsänderungen (Volt)

Wandlerprinzip:

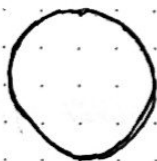
- Passiv: Akustische Energie wird direkt in elektrische gewandelt
- Aktiv: Verändern eine zugeführte elektr. Energie im Rhythmus der Schallschwingungen

Kondensator-Mikrofon: Signal ist proportional zur Membranauslenkung

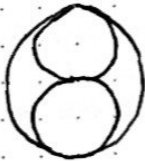
↳ Anlegen einer elektr. Spannung führt zu Potenzialänderung zw. Membran und Platte
 $\Rightarrow$ Spannungsschwankungen = elektr. Signal

Bauformen:

- Kugel



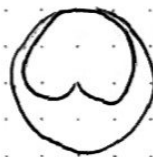
- Acht



- Kegel



- Niere



Sensor: Bauteil, das chem./phys. Eigenschaften seiner Umgebung qualitativ oder quantitativ erfassen kann

ERFASSEN:

- Aufnahme von anal. Signalen
- Wandlung in dig. Messdaten
- Verarbeitung durch Software

MESSEN: Experimentelle Bestimmung des Messwertes einer physikalischen Größe durch einen quantitativen Vergleich der Messgröße mit einer Einheit als Bezugsgröße

Messgröße: diejenige phys. Größe, der eine Messung dient

Messwert: Vom Messgerät gelieferte Wert einer Messgröße

Phys. Größe sind z.B. Länge, Stromstärke
Einheiten sind dann z.B. Meter, Ampere

Messfehler: Abweichung des Messwertes von der Messgröße

TON: Signal aus einer Frequenz

KLANG: Mehr als eine Frequenz

GERÄUSCH: Viele unterschiedliche Frequenzen

SPRACHE: Wie Geräusch, nur dass Menschen diese mit erlernten sprachlichen Wissen kombinieren.

Es gibt zwei Ebenen in der Entwicklung von Sensordatenverarbeitungssystemen:

1. ENTWURFSEBENE: Von der anschaulichen Aufforderung zur mathematischen Spezifikation
2. IMPLEMENTIERUNGSEBENE: Von der mathematischen Spezifikation zu Rechnerverhalten / Code

Beispiele:

BILD:

- Qualitätskontrolle
- Fahrerassistenz
- Photogrammetrie
- 3D Modell aus Bildern
- Objekterkennung

TON:

- Digitale Assistenten
- Hands-free wearables
- Transcriptions
- Language barriers

BEWEGUNG:

- Wearable Computing
- Airwriting
- Gimballs
- Motion Capturing

SDV4: Farbe

Licht ist eine elektromagnetische Welle. Sichtbar von 380 - 720 nm.

Die Farbe eines Objekts sind die Wellenlängen, die nicht verschluckt werden.
Farbige Bildsensoren im Mosaikmuster vor einem Monochromen Image Sensor

R	G
G	B

 ← weil Menschen für Grün empfindlicher sind.

Segmentierungsbetriebene BV:

Segmentierung → Gruppierung → Kennzahlen → Filtern & klassifizieren

- ↳ Klassifikation einzelner Pixel durch Grauwert Schwellwert
- ↳ Farbsegmentierung: Pixel klassifizieren anhand gegebener Bereiche der HSV Skala
 - ↳ Oder: Hintergrund erkennen und invertieren

↓
Hue = Farbe
Saturation = Sättigung
Value = Helligkeit

Kantenerkennung mit Canny-Algorithmus

Gruppierung: zu einem Objekt gehörende Pixel zusammenfassen

- ↳ Probleme:
 - 1) Objekte berühren sich und werden zu einem
 - 2) Objekt zerfällt durch Erkennungsfehler in mehrere Objekte

- ↳ Lösungen:
 - 1) Wasserscheidentransformation:
 - Objektkerne durch aggressive Erosion bestimmen
 - Regionen nach Objektkernen berechnen
 - Dehne Regionen aus (Dilatation)
 - 2) Morphologie: Bearbeiten der binarisierten Bilder
 - Opening / Closing

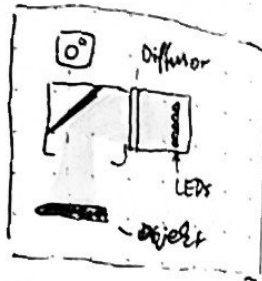
SDV3: Datenaufnahme Bild

Lichtstrahlen $\Rightarrow$ Elektrisches Signal $\Rightarrow$ Digitales Signal
└ Verdeckung, Schatten, Reflexionen, Indirektes Licht, Beleuchtungsgradient

Diffusor verteilt Licht gleichmäßig

Koaxialbeleuchtung: Halbkugliger Spiegel teilt Strahlen zwischen Kamera und Licht

Dunkelfeldbeleuchtung: Licht im extrem flachen Winkel, nur Erhebungen reflektieren in die Kamera



(Dunkelfeld) - Durchlicht: Licht kommt von hinter dem Objekt

Polarisation: Licht in Wellenform (sowohl links $\leftrightarrow$ rechts, als auch oben $\leftrightarrow$ unten), mit Polarisationsfilter (vor Lichtquelle und Kamera) lassen sich Wellen einer Richtung blockieren

Bild- / Brennweite: Bildausschnitt verändern, Verzerrungen ab $\sim 45^\circ$

Linse bricht viel Licht auf einen Punkt. Dadurch helles Ergebnis

└ Faktoren für Unschärfe: - Entfernung des Objektes
- Linsendurchmesser

Bereich akzeptierter Schärfe $\Rightarrow$ Schärfentiefe

Kleine Blendenöffnung = große Blendenzahl $\Rightarrow$ Große Schärfentiefe
Große Blendenöffnung = kleine Blendenzahl $\Rightarrow$ Kleine Schärfentiefe

Bildsensoren im Rasterformat angeordnet

Bewegungsunschärfe durch Belichtungszeit $\Rightarrow$ Hohe BLZ für dunkle Umgebung

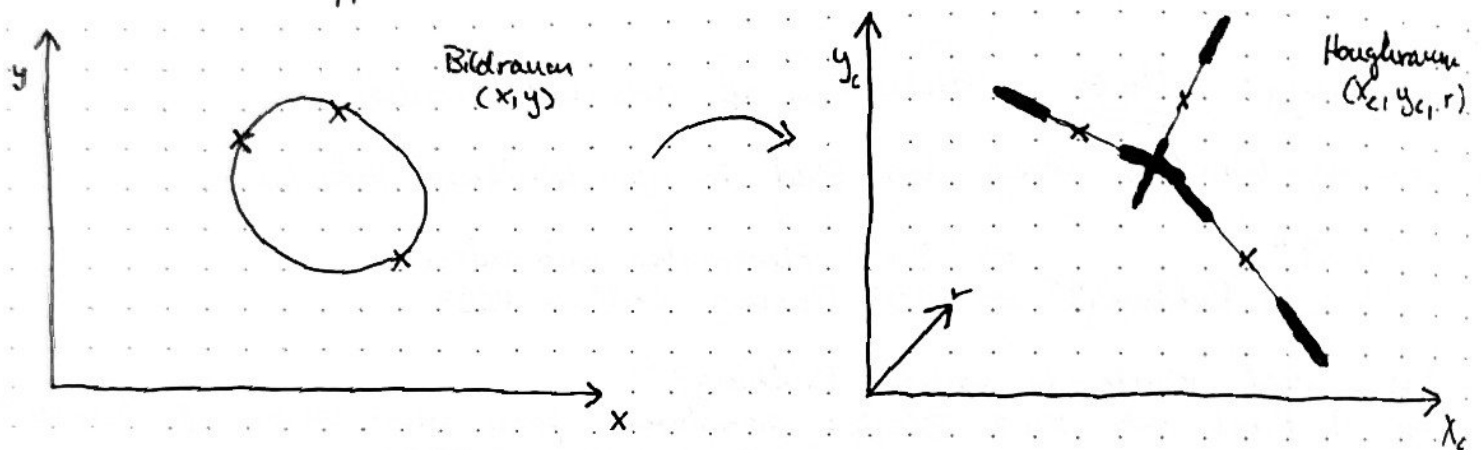
Shutter: - Global = Alle Pixel belichten synchron
- Rolling = Pixel starten zeitversetzt.
 $\Rightarrow$ Unterschiedliche Zeiten zeigen Realität unterschiedlicher Zeitpunkte

SDV9: Houghtransformationen, Bewegungsmerkmale

Hough-Transformation zur Erkennung von Kurven / Kreisen

- Akkumuliert Evidenz im Houghraum
- 3D-Houghraum kostet Speicher und Rechenzeit
 - ⇒ 2D-Houghraum mit Sobelrichtung in x und y bei jedem erkannten Kantenpunkt im Bildraum
 - ⇒ Sobelrichtung mit $r_{\min}$ und $r_{\max}$
 - ⇒ Kreismittelpunkt durch höchste Konzentration von Sobelvestoren

Problem: Überlappende Kreise werden schlecht erkannt



Invariante Merkmale ändern sich nicht

Kovariante Merkmale unterliegen der selben Transformation

Merkmale im Zeitbereich:

- Energie, Area under Curve
- Maximum, Minimum, Steigung
- Zero-Crossing Rate (wie oft $x=0$?)

Stat. Verteilungsmerkmale:

- Mittelwert, Standardabweichung, Varianz
- Maximum, Minimum
- Quadratisches Mittel

SDV8: Kanten, SDV-Paradigmen, direkte Bildmerkmale

Kantenerkennung = 3×3 -Feld um Pixel betrachten

↳ Sobelfilter = $\begin{matrix} x & \text{vertikale Kanten} \\ y & \text{horizontale Kanten} \end{matrix}$

↳ Sobellänge = Proportional zur Intensität der Kante
↳ Sobelwinkel = Winkel quer (90°) zur Kante

SDV-Paradigma: Keine frühen Entscheidungen!

Kantenerkennung: HOG (Histogram of Oriented Gradients)

↳ Sliding-Window über das Bild in verschiedenen Größen

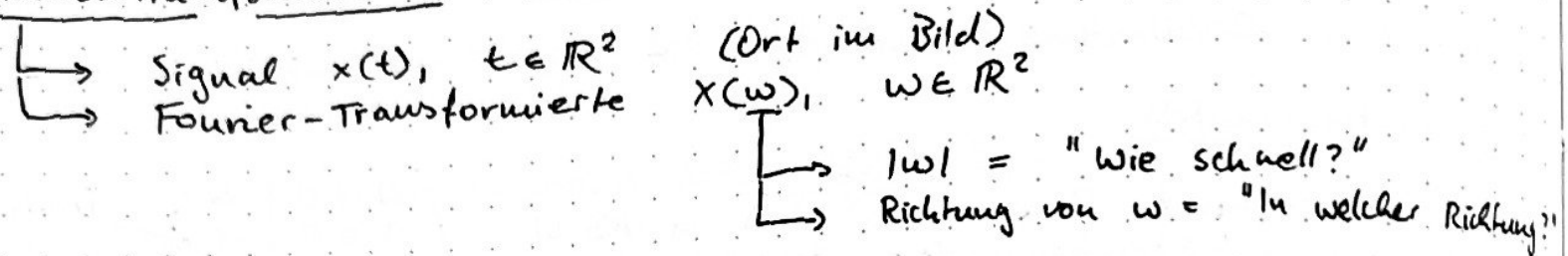
HOG = "Wo?" $\Rightarrow 8 \times 8$ Pixelraster betrachten
"Welche Richtung?" $\Rightarrow 20^\circ$ Raster ($\times 9$) = 180°

↳ "Wie viel Kante in welche Richtung?"
 $\Rightarrow$ 1 Pixel aus 8×8 Block: Sobelwert passt nicht genau auf Raster?
 $\Rightarrow$ Anteilig verteilen

2×2 HOG Raster auf Länge 1 normieren
 $\Rightarrow$ Überlappende 2×2 Blöcke für Normalisierung

SDV 7: 2D-Frequenzraum, Filter

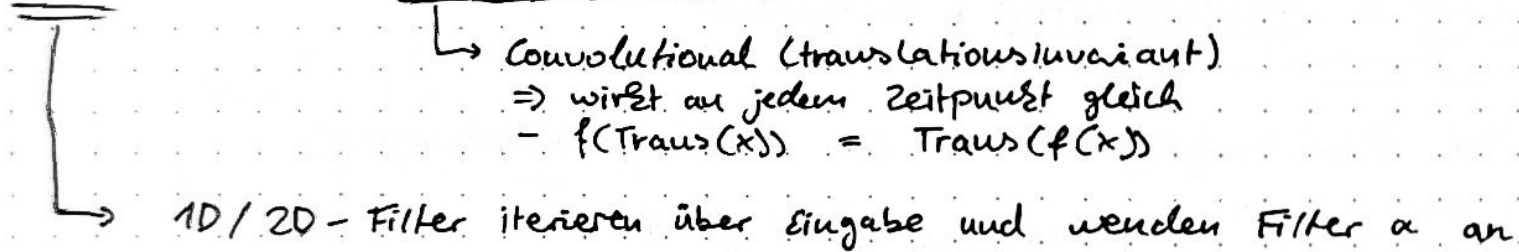
2D-Fouriertransformation:



⇒ Zerlegt 2D-Signal $x(t)$ als Summe von Sinuswellen verschiedener Frequenzen und Richtungen

Nutzen von 2D-FT: z.B. JPEG-Kodierung, laut Trennung von Groben und Feinen

Filter: Lineare, translationsinvariante Abb. von Signalen, oft lokal



Filtern: Multiplikation im Frequenzraum

⇒ Verstärkung und Phasenverschiebung je nach Frequenz

$$\Rightarrow F(a * x) = F(a) F(x)$$

Diagram illustrating the relationship between convolution and Fourier transformation:

- ↑ Konvolution
- ↑ Fouriertransformation

SDVG: Koordinatensysteme, Inertialsensor

Position + Orientierung eines Objekts ergibt Pose

Gegeben zwei Koordinatensysteme $A, W = A_x^{(W)} \leftarrow x\text{-Rotation von } A \text{ in } W$

Transformationsmatrix

- Transformiert Pose eines ^{Objekts aus} KS A in KS B
- Aufbau / Struktur der Matrix:
- $T_{C \leftarrow A} = T_{C \leftarrow B} \cdot T_{B \leftarrow A}$
- $T_{B \leftarrow A} = (T_{A \leftarrow B})^{-1}$

Orientierung $R_{B \leftarrow A}$

$$\left(\begin{array}{ccc|c} 0 & -1 & 0 & 2 \\ 1 & 0 & 0 & -1 \\ 0 & 0 & 1 & 0 \\ \hline 0 & 0 & 0 & 1 \end{array} \right) \begin{array}{l} \text{Position} \\ \leftarrow B \leftarrow A \\ \\ \\ \text{Rest (fest)} \end{array}$$

Inertialsensor = Accelerometer + Gyrometer

Misst Beschleunigung in 3 Achsen (MINUS Erdbeschleunigung)

$$\Rightarrow V_{t+dt}^{(W)} = V_t^{(W)} + dt(R_{W \leftarrow t} a_t^{(IE)} + g^{(W)})$$

$$\Rightarrow P_{t+dt}^{(W)} = P_t^{(W)} + dt V_t^{(W)}$$

- $R_{W \leftarrow t}$ = Orientierung Inertialsensor in Welt
- $g^{(W)}$ = Gravitationsvektor
- $a_t^{(IE)}$ = Beschleunigung (Accelerometer-Messung)

Gyrometer misst Drehungen $\Rightarrow$ Licht durchläuft Ring in beide Richtungen
Bei Bewegungen kürzerer Weg $\Rightarrow$ wird gemessen

$$\Rightarrow R_{W \leftarrow t+dt} = R_{W \leftarrow t} \underbrace{\text{Rot}(dt \omega_t^{(IE)})}_{R_{t \leftarrow t+dt}} \quad \left| \quad \omega_t^{(IE)} = \text{Drehrate (Gyrometermessung)} \right.$$

Beim Inertialsensor akkumulieren sich Fehler:

- Gyrofehler
- Orientierungsfehler $\rightarrow$ Gravitationsfehler
- Geschwindigkeitsfehler
- Positionsfehler

$\Rightarrow$ Dadurch ist Inertialsensor KEIN dauerhafter Posensensor

$\Rightarrow$ ABER dauerhafter Neigungssensor // Orientierungssensor

SDVS: 1D-Frequenzraum, Fouriertransformation

Mikrofon: Schall in elektrisches Signal gewandelt = analoges Signal

- ↳ Samplingrate: Wie oft pro Sekunde
- ↳ Bittiefe: Wie viele Stellen

Wandlung von analoges Signal $\Rightarrow$ digitales Signal: 1) Abtastung
(zeitkontinuierlich $\rightarrow$ zeitdiskret)

Analysen von Audio oft im Frequenzbereich
 $\Rightarrow$ Fouriertransformation

2) Quantisierung
(wertkontinuierlich $\Rightarrow$ wertdiskret)

Fouriertransformation: Signal $\Rightarrow$ Frequenzspektrum

- ↳ lässt sich invertieren: keine Informationen werden verloren
- ↳ Interessant für Spracherkennung: DTFT (Discrete-Time Fourier transform)

DTFT: - Eingabe: Diskret, endlich, aperiodisch
- Frequenzspektrum: Kontinuierlich

- Normiert Frequenzen des dig. Signals auf Bereich $[-\pi, \pi]$
- Max Frequenz π im dig. Signal $\leftrightarrow$ Nyquist-Frequenz im kont. Signal
- In der DTFT sind nur Frequenzen unterhalb der halben Abtastrate (Nyg. Freq.) sonnt
- Diskrete Frequenzen π entsprechen der kontinuierlichen Nyquist Frequenz
- Algorithmus: Fast Fourier Transform (FFT)

Spektrogramm stellt Zusammensetzung eines Signals aus einzelnen Frequenzen im zeitlichen Verlauf dar

↳ STDTF = (short-time discrete Fourier transform) = FT auf best. Bereich

Weitere Bsp: - Maschinenakustik
- 2D-sonogramme, Ultraschall

SDV12: Abstraktionsebenen für SDV

Anforderungen

- Was soll das System können?

F: Kunde braucht eigentlich etwas anderes

Entwurf

- Abfolge von Algorithmen

F: Verfahren leistet nicht das Intendierte

Parameter

- Schwellwerte für die Algorithmen

F: Werte sind für gegebene Situation nicht korrekt

Implementierung

- Quellcode

F: Bug

← Häufigste Fehlerquelle

Metriken zur Auswertung:

- precision: $\frac{tp}{tp+fp}$

- recall: $\frac{tp}{tp+fn}$

- accuracy: $\frac{tp+tn}{tp+fp+fn}$

		true	
		p	n
test	p	tp	fp
	n	fn	tn

SDV11: Klassifikation

Klassifikation: Zusammenfassen von Objekten zu Klassen

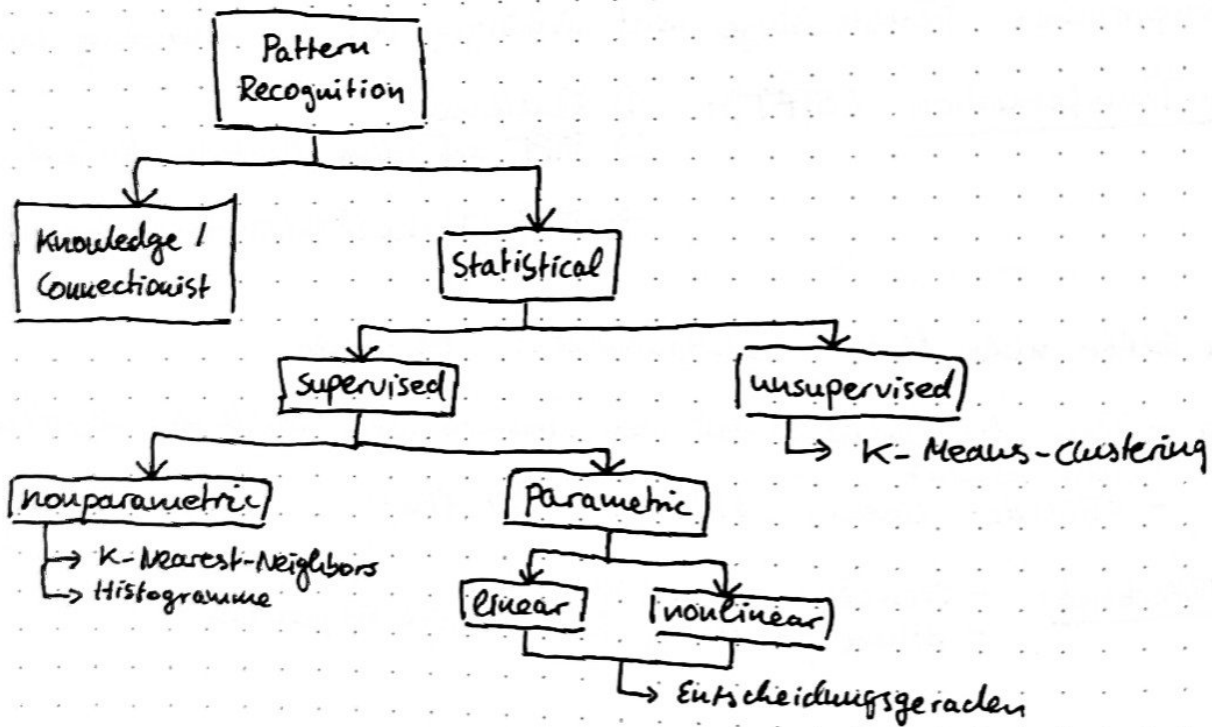
- Token-Kl.: Einzelne Merkmalsvektoren werden einer Klasse zugeordnet
- Sequenzen-Kl.: Sequenzen von Merkmalsvektoren werden in ihrem Kontext zugeordnet

Klassifikator: Ordnet Merkmalsvektoren in Klassen ein

- Meist ein Algorithmus oder eine mathematische Funktion

Zweidimensionaler Merkmalsraum

- Entscheidungsgerade (linear oder nichtlinear)
 - Herausforderung: Passende Anzahl an Parametern
 - ⇒ Overfitting / underfitting
- K-Nearest-Neighbors:
 - Betrachte die K nächsten Nachbarpunkte und klassifiziere entsprechend



SDV10: Audiomerkmale

Merkmale sind erkennbare Eigenschaften, die einen abstrakten Zusammenhang von anderen unterscheidet

- z.B. Farbe, Größe, etc.
- Bei Auswahl von Merkmalen auf Komplexität achten (weniger ist mehr)

Bsp: "KO" von "OK" voneinander unterscheiden:

- Im Spektrogramm sind deutliche Unterschiede erkennbar
- Im DTFT allerdings nicht: DTFT analysiert welche Frequenzen enthalten sind, aber nicht zu welchem Zeitpunkt

⇒ SDTFT: Partitionieren des Signals in eine Sequenz kürzerer Segmente (windowing)

- Dabei überlappen sich Segmente (gewollt)
- Fensterfunktion: Rechteckfenster, Hanningfenster
- Zu kurze Fenster: Ungenau im Frequenzbereich
- Zu lange Fenster: Ungenau im Zeitbereich

⇒ Kompromiss: Fensterlänge wird abhängig von der Anwendung festgelegt

Kurzzeit-Fourier-Transformation (STFT): 1) Fensterung
2) DTFT auf jedem gefensterten Abschnitt (Frame)

→ DTFT gibt eine Frequenzverteilung für jeden Frame

Sprache wird im Gehör nicht linear wahrgenommen: Mel-Skala

⇒ Mel-Skala: - Ton, der doppelt so hoch wahrgenommen wird, erhält den doppelten Tonheitswert
- < 500 Hz: Linear, > 500 Hz: Nicht linear

⇒ Mel-scale-Filterbank: - Dreieckige Gewichtungsfunktion
- Höhere Auflösung für niedrige Frequenzen