

Onset Detection

Was ist Musik?

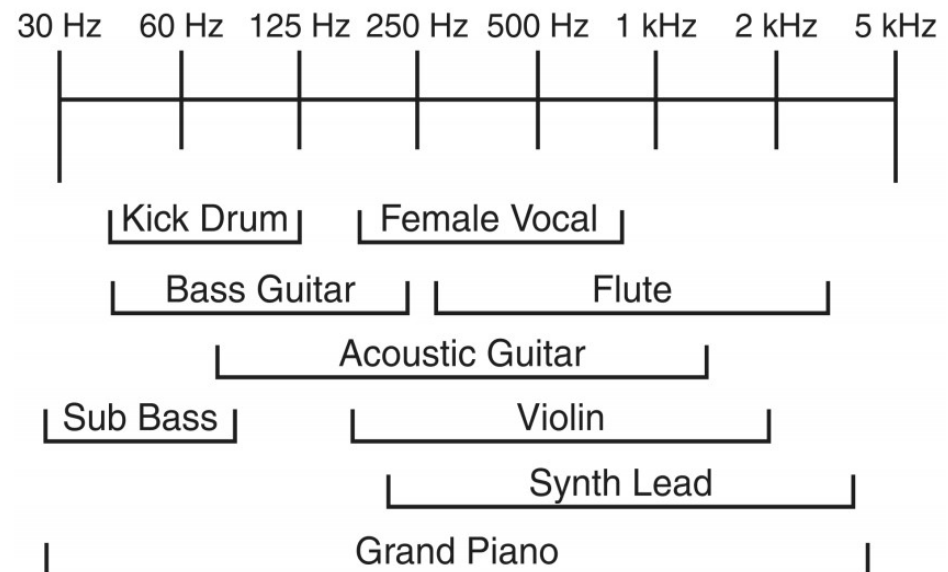
- Gute Frage. Nächste Frage?
- Features:
 - Tonhöhe
 - Klangfarbe
 - Lautstärke
 - Rhythmus
 - Tempo
 - Takt

Töne

- Die meiste Musik besteht aus Tönen und Geräuschen
- Parameter
 - Tonhöhe
 - Intensität
 - Klangfarbe

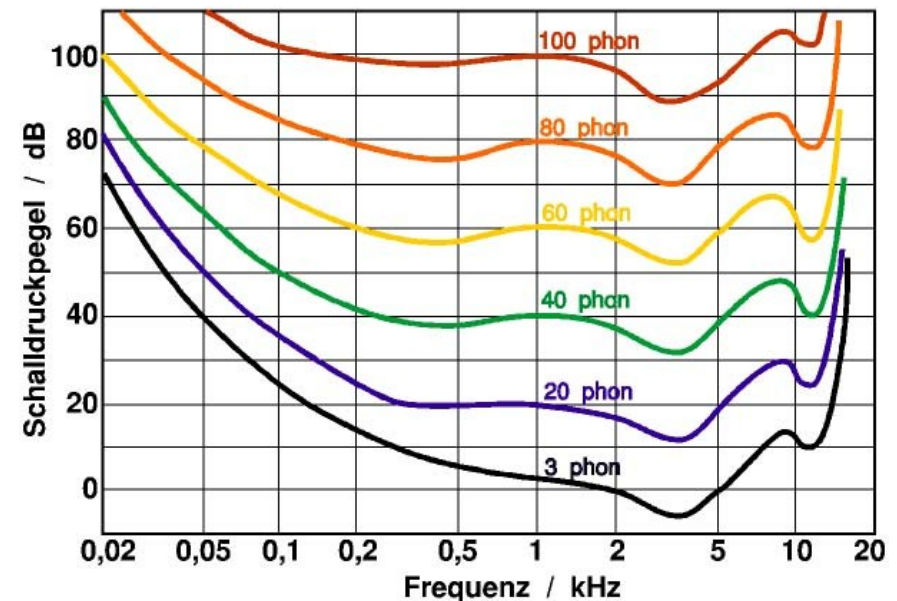
Tonhöhe

- Die Frequenz eines Tones
- Bestimmt die Note
- Hörvermögen des Menschen: 20Hz – 20kHz
- Frequenzen die in Musik direkt verwendet werden: 30Hz – 5kHz
- Frequenzband wird in Oktaven unterteilt
- 12 Töne je Oktave



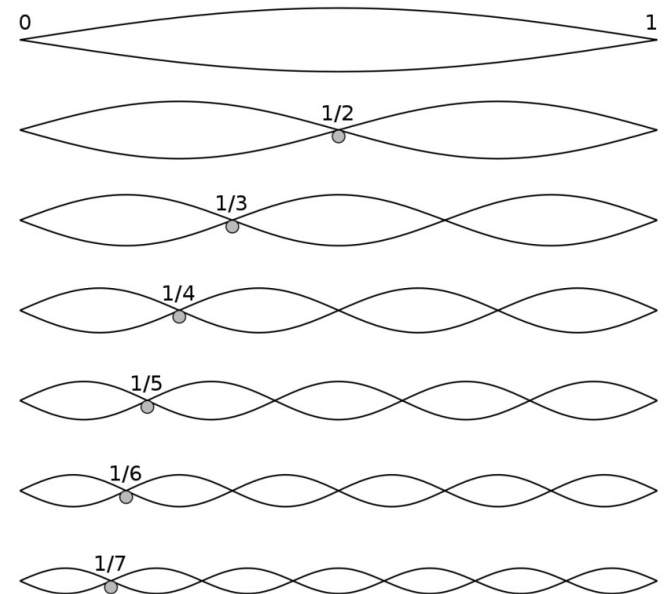
Intensität

- oder Amplitude
- Einheit: dB
 - 0dB: das leiseste was der Mensch hören kann
 - 70dB: eine Unterhaltung
 - 120dB: ein startendes Flugzeug aus 60m
- Amplitude \neq Lautstärke



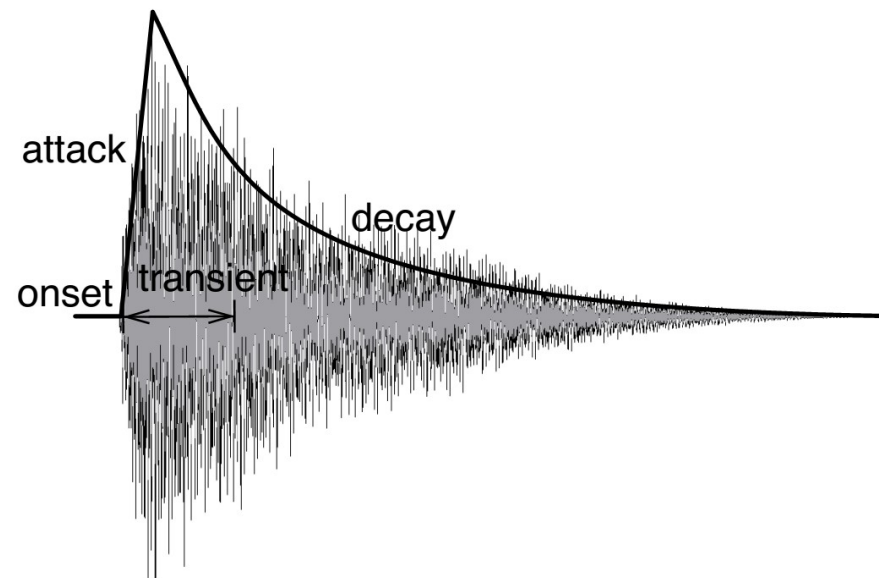
Klangfarbe

- Töne klingen je nach Instrument unterschiedlich
- Ton ist aus Grundton und Obertönen zusammengesetzt
- Obertöne können harmonisch sein



Was sind Onsets?

- Onset: der Start einer Note oder eines Geräusches
- Attack: die Amplitude steigt von 0 bis zum Maximum
- Decay: das Signal klingt wieder ab bis es 0 ist



Was sind Onsets?

- Pitched Percussive: Klavier, Gitarre, ...
- Pitched Non-Percussive: Geige, Gesang, ...
- Non-Pitched Percussive: Becken, Drums, ...

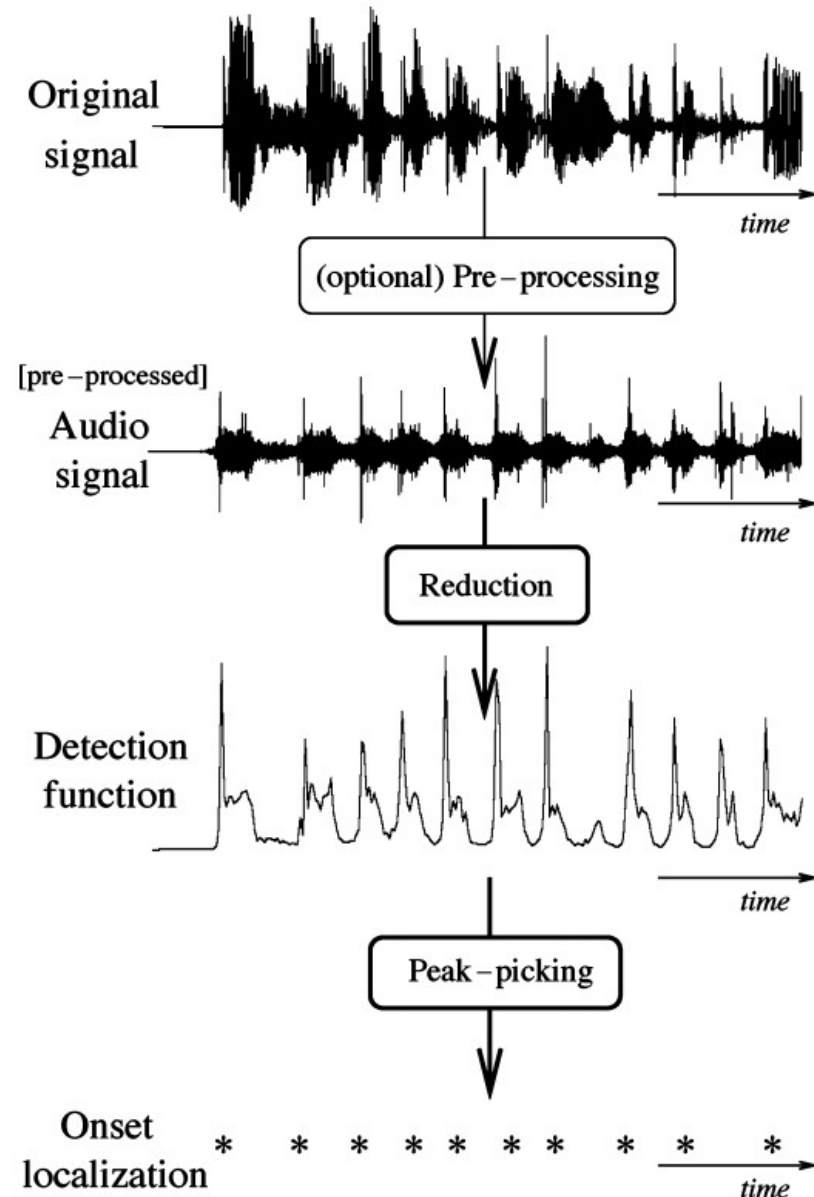


Onset Detection ist...

- ein erster Schritt für die Feature Extraktion von Audio-Signalen / Musik
- die Basis für viele Techniken in der Audio- bzw. Musikanalyse (Beat-, Tempo-Extraktion, ...)
- nicht trivial:
 - Signale haben meist schlechte Qualität (Rauschen, ...)
 - Polyphone Musik
 - Unterschiedliche Arten von Onsets
 - Onsets überlappen sich

Onset Detection Algorithm

- Ablauf der Onset Detektion:
 - Pre-Processing
 - Reduktion auf Detection Function
 - Peak-Picking



Pre-Processing

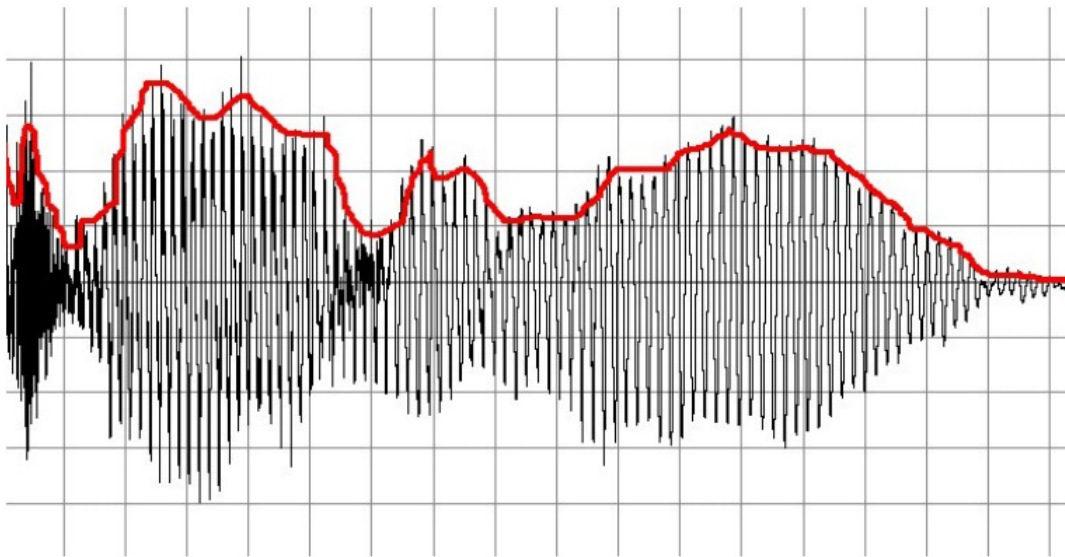
- Ziel
 - Aspekte des Signals verstärken oder abschwächen
 - Für Reduktion aufbereiten
- Methoden
 - FFT, STFT, Wavelet-Transformation
 - Rauschentfernung
 - Signal in Frequenzbänder aufteilen
 - ...

Reduction

- Ziel
 - Signal in unterabgetastete Detection Function transformieren
 - Detection Function soll transienten hervorheben
- Methoden basieren auf
 - Steigung der Hüllkurve
 - Spektrale Features
 - Phase des Signals

Hüllkurve

- Sehr einfacher Ansatz
 - Nach Onsets kommt ein starker Anstieg der Amplitude
 - Dafür (geglättete) Hüllkurve verwenden



$$E_0(n) = \frac{1}{N} \sum_{m=-\frac{N}{2}}^{\frac{N}{2}-1} |x(n+m)| w(m)$$

$$E_0(n) = \frac{1}{N} \sum_{m=-\frac{N}{2}}^{\frac{N}{2}-1} [x(n+m)]^2 w(m)$$

w ist ein Fenster mit N
Punkten und um m=0
Zentriert

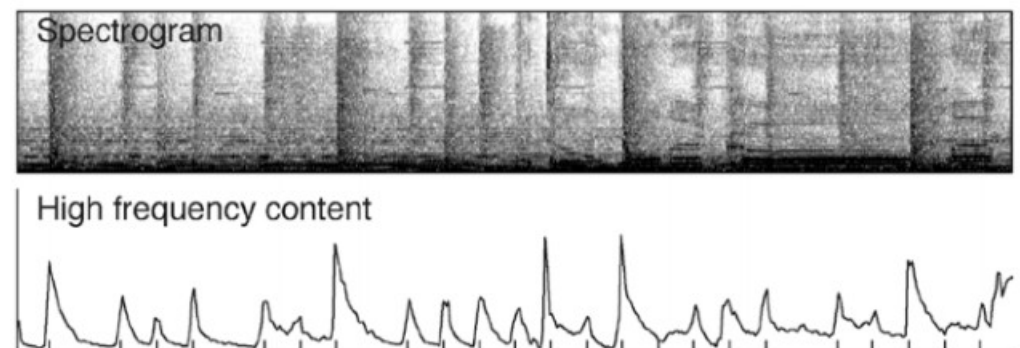
HFC (High Frequency Content)

- Ansatz basiert auf
 - meist sehr breitbandigen Transienten
 - Energie des Signals bei niedrigen Frequenzen (30Hz – 5kHz)
 - Stärkere Gewichtung von hohen Frequenzen

$$\tilde{E}(n) = \frac{1}{N} \sum_{k=1}^N W_k |X_k(n)|^2$$

$X_k(n)$ ist die Energie im Band k der FFT zum Zeitpunkt n

W_k ist die Gewichtung
z.B. linear $W_k = |k|$



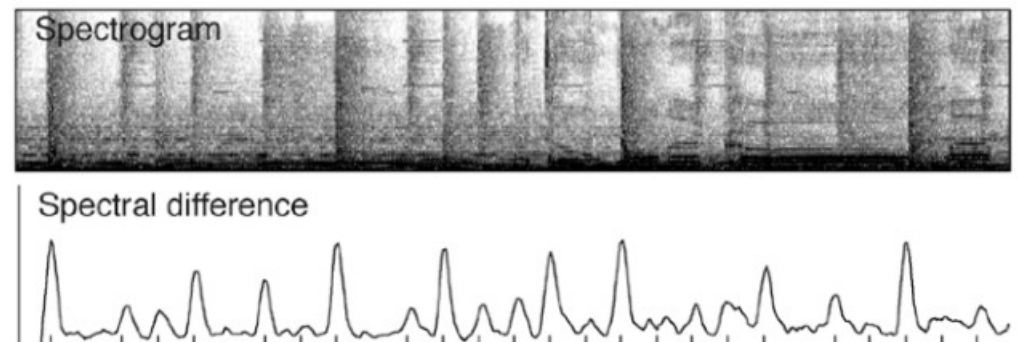
Spectral Flux

- Ansatz basiert auf
 - Zeitlicher Entwicklung des Signals
 - Der Veränderung im Spektrum zu 2 Zeitpunkten

$$SD(n) = \frac{1}{N} \sum_{k=1}^N \{ H(|X_k(n)| - |X_k(n-1)|) \}^2$$

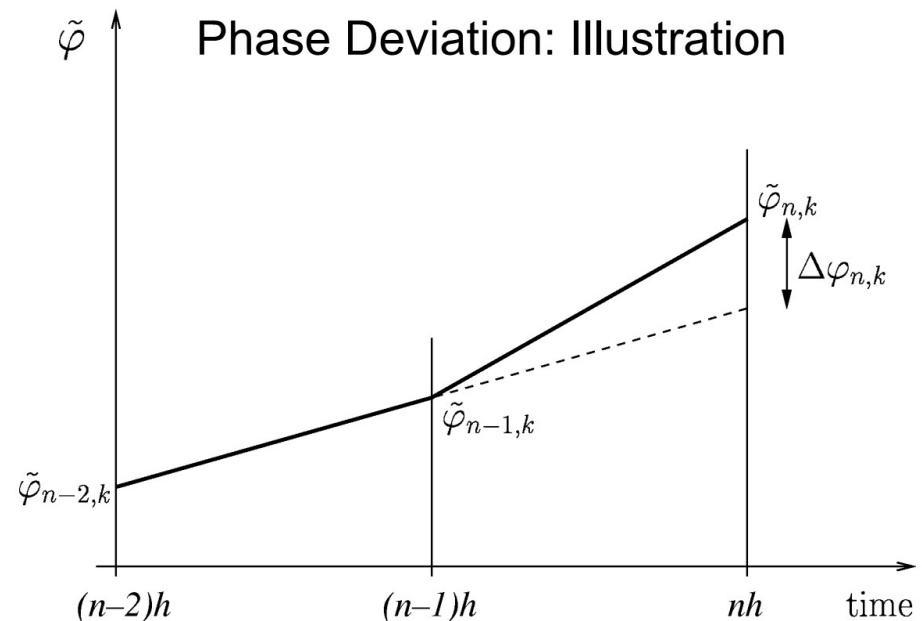
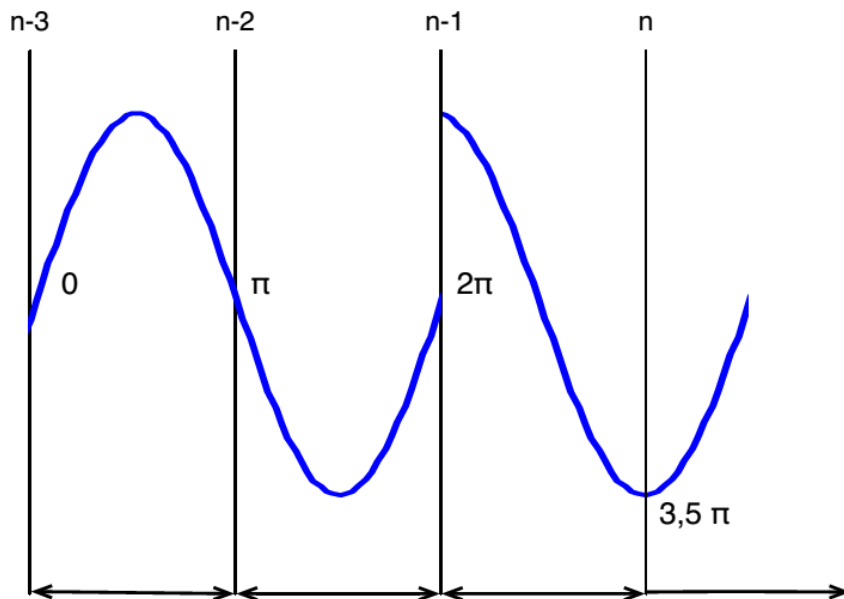
$$H(x) = (x + |x|) / 2$$

$H(x)$ bewirkt dass nur Anstiege der Energie mit einbezogen werden



Phase Deviation

- Ansatz basiert auf
 - Dem Wissen das ein Sinus in einem Fenster immer die gleiche Phase hat
 - Transienten eine Änderung der Phase bewirken

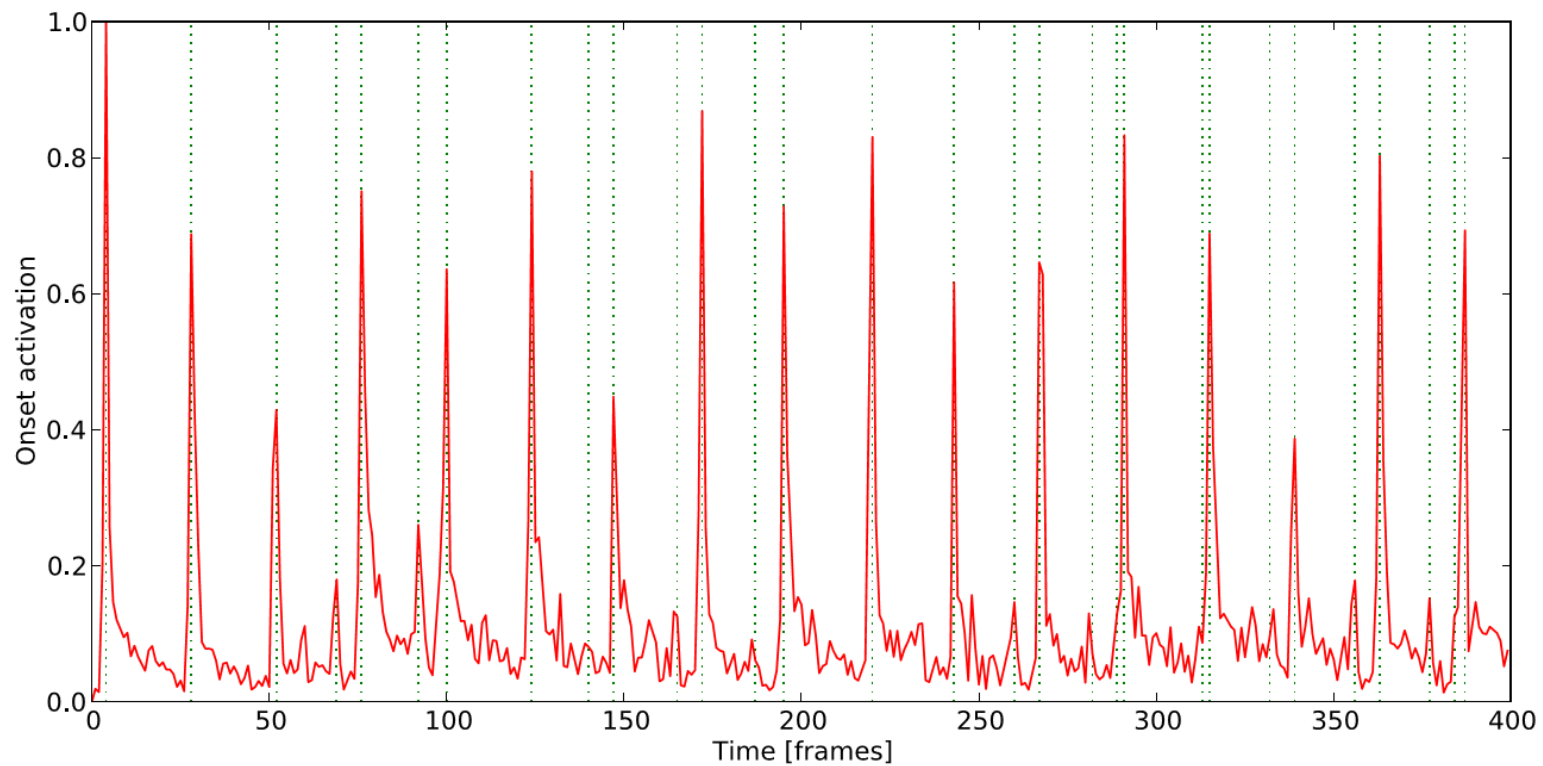


Peak Picking

- Ziel
 - Die Onsets anhand der Detection Function identifizieren
 - Maxima in Detection Function finden
- 3 Schritte nötig
 - Post Processing
 - Thresholding
 - Entscheiden

Peak Picking

- Detection Function

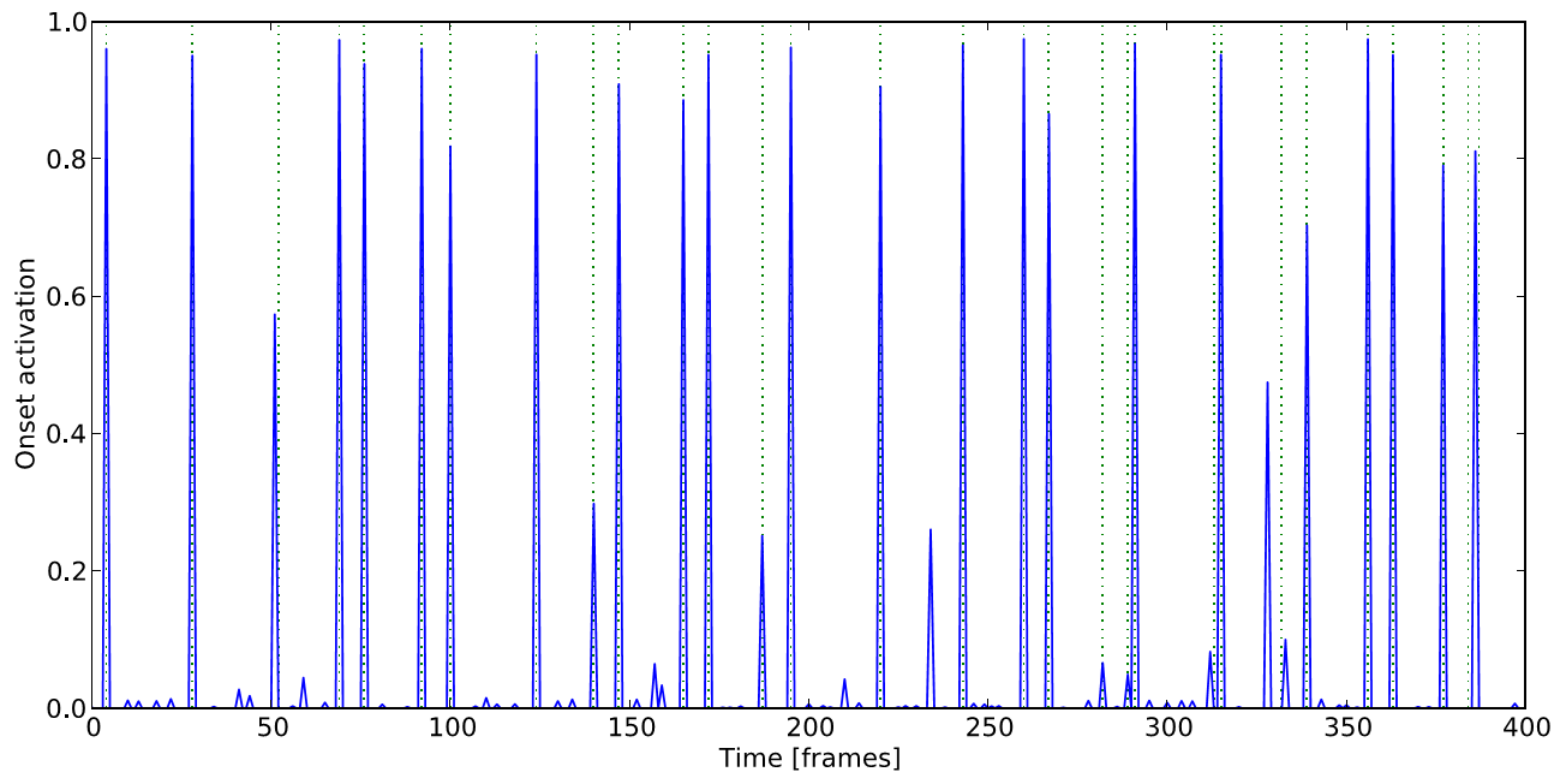


Post Processing

- Zweck
 - Einheitlichkeit und Konsistenz der Features in Detection Functions erhöhen
 - Onsets in isolierte, leicht nachweisbare, maxima transformieren
- Methoden
 - Rauschentfernung
 - Normalisieren
 - DC Anteil entfernen

Post Processing

- Detection Function
 - Nach Post Processing

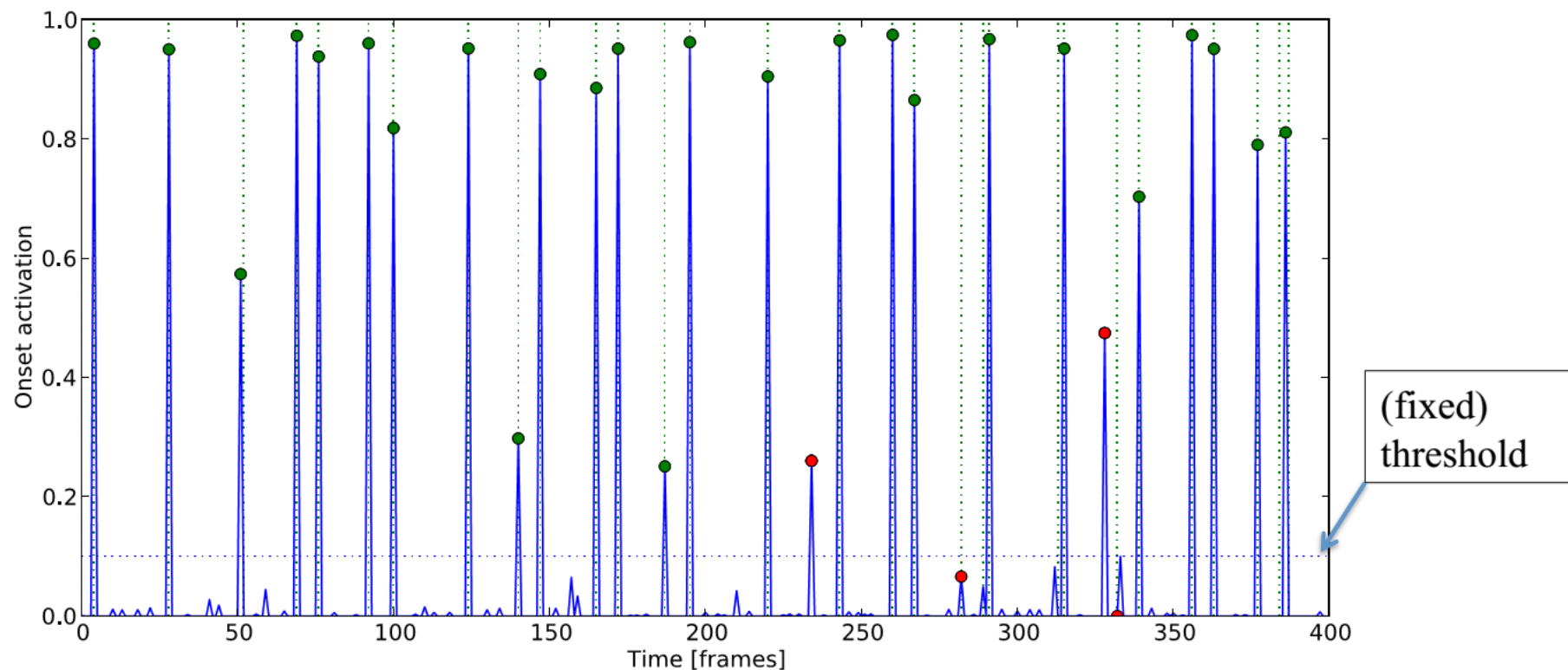


Thresholding

- Es gibt Maxima die keine Onsets sind
- Nur Maxima über einer Schwelle als Onsets auswählen
- Fixe Schwelle ist nicht optimal
 - Nur für Signale mit wenig Dynamik
 - Erkennt Onsets wenn Maxima über bestimmten Wert sind
- Adaptive Schwelle besser
 - Schwelle ist stark gefilterte Detection Function (TP)
 - Hohe Maxima überdecken niedrigere

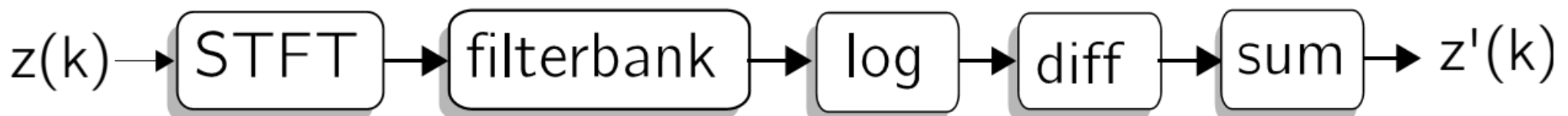
Thresholding

- Detection Function
 - Mit Schwellwert
 - Mit markierten Onsets



'SuperFlux'

- State of the art Algorithmus
 - Basierend auf Spektral Difference
 - Verwendet Psychoakustische Effekte
 - Optimiertes Peak Picking



'SuperFlux'

- Berechne STFT
 - window size = 23ms, hop size = 10ms
- Filtern
 - Spektrum auf Semitone-Skala abbilden
- Logarithmische Amplitude
 - Logarithmus auf Amplitude anwenden
- Spectral Flux
 - Spectral Flux als Detection Function berechnen

'SuperFlux'

- Optimiertes Peak Picking
 - Jedes Maximum muss folgende Kriterien erfüllen
 - Es muss das Maximum in einem Fenster sein
 - Es muss größer als der Mittelwert in einem Fenster sein
 - Eine Mindestzeit zum letzten Onset darf nicht unterschritten werden