Transient Detection With Absolute Discrete Group Delay

Volker Gnann, Martin Spiertz

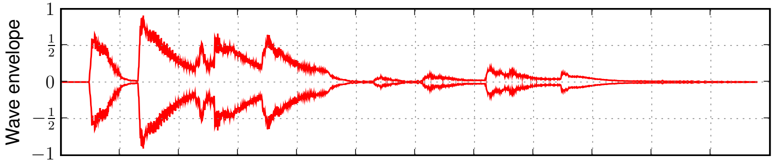
Institut für Nachrichtentechnik
RWTH Aachen University

Overview

- Transient Detection
- The Average Absolute Discrete Group Delay (AAGD)
- Metafiltering: Why the AAGD criterion works
- Improvements
- Performance

Transient Detection

Transient = quickly and un-predictably changing audio signal segment
Onset = starting point of a note



Current STFT-based Methods

- Energy change
- Phase propagation over time (pitch change)

Disadvantage of both

Knowledge of neighboring frames required

New Approach

Phase coincidence over frequency within an STFT time frame

Review: Short-Time Fourier Transformation

STFT Definition

$$X(m,k) = \sum_{n=-\infty}^{\infty} x(n)w(mS-n)\mathrm{e}^{-j\frac{2\pi k}{N}n}$$

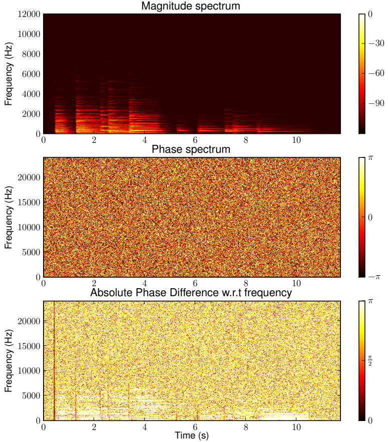
Results are complex and consist of

- magnitude
- energy-independent phase

Phase is ambiguous

- 2π = full angle
- Restriction to $[-\pi, \cdots, +\pi]$ range

$$\mathrm{princ}(\varphi) := ((\varphi + \pi) \bmod (-2\pi)) + \pi.$$



V. Gnann, M. Spiertz   Institut für Nachrichtentechnik   RWTH Aachen University

Average Absolute Group Delay

Idea: Use the phase evolution **across frequency** to detect transients.

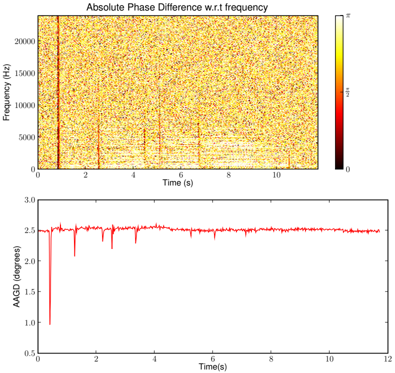
Definition AAGD

$$D_X(m,k) = \mathrm{princ}(\varphi_X(m,k) - \varphi_X(m,k-1))$$

$$AAGD_X(m) = \frac{1}{N}\sum_{k=1}^{N} |D_X(m,k)|$$

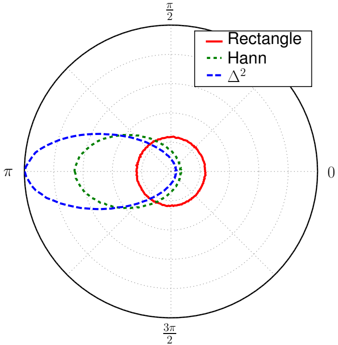
Transient Detection

- Learn mean μ and variance σ from AAGD
- Thresholding: $\theta = \mu - \lambda\sigma$
- λ determines sensitivity (≈ 2)



Distribution of phase differences

Pink Noise:

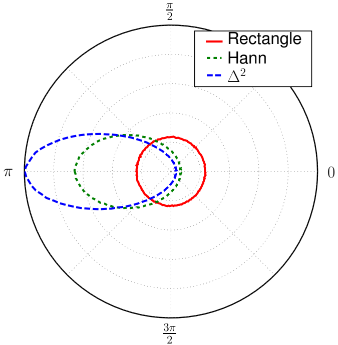


Conclusion for steady-state signals

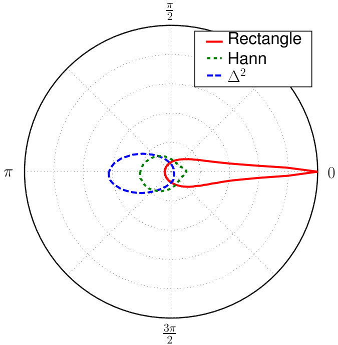
$AAGD \longrightarrow \pi$ for bell-shaped windows. Reason: next pages

Distribution of phase differences

Pink Noise:



Oboe:



Conclusion for steady-state signals

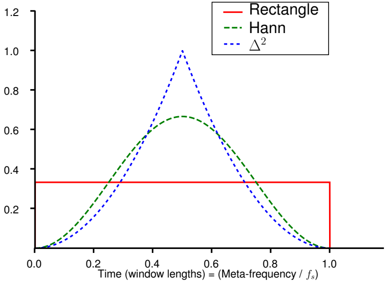
$AAGD \longrightarrow \pi$ for bell-shaped windows. Reason: next pages

Definition

- Windowing in time domain = convolution in frequency domain
  $x(n) \cdot w(n) \xrightarrow{\mathcal{F}} X(k) * W(k)$
- Convolution in frequency domain = FIR "filtering" the DFT coefficients (Meta-Filtering)

Properties

- IDFT = time-reversed DFT
- Symmetric signal: IDFT = DFT
- Transfer function of meta filter = window form in time domain
- Bell-shaped window $\Longrightarrow$ high-pass meta filter

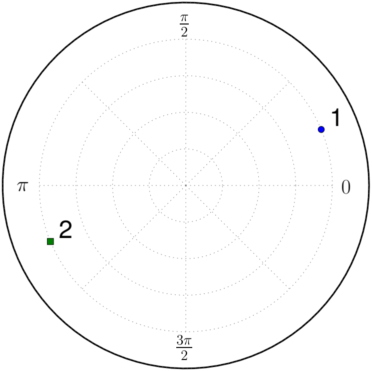


Consequences of Metafiltering

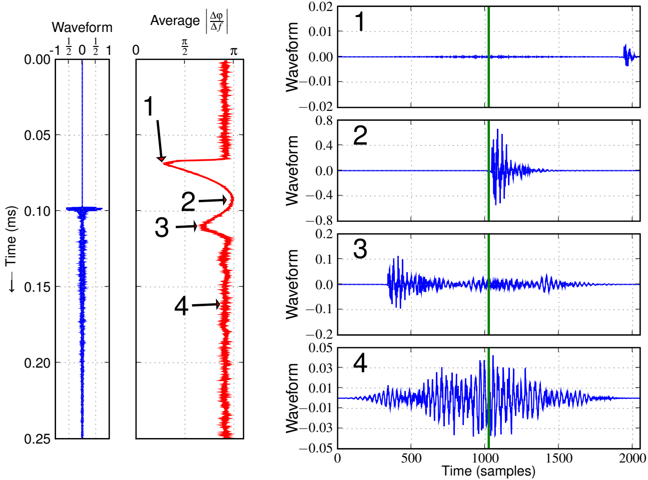
Metafiltering and Phase

- High-pass meta filter = amplification at Nyquist frequency
- strong oscillation between neighboring frequency coefficients

  $\Longrightarrow$ alternating phases



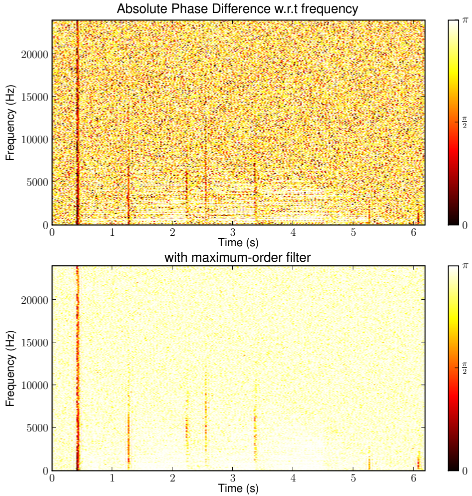
Transient Behavior: A Castanet Clap



V. Gnann, M. Spiertz Institut für Nachrichtentechnik RWTH Aachen University

Improvements

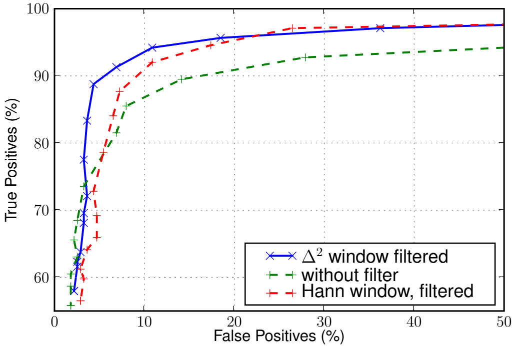
- Denoising with a maximum-order filter
- Restricting the frequency range
- Adding noise (to overwrite note-off events)



Onset Detections of Percussive Sounds

Test Set

- Source: EBU-SQAM
- percussive and tonal-percussive instruments
- 276 hand-marked reference onsets in 43 files
- Tolerance: 50ms before transient, 0 ms after
- Added noise: white uniformly distributed, -34 dB peak level



Summary

- New transient detection method based on average absolute group delay
- Detects percussive and tonal-percussive sounds
- Influence of window function:
  - Convolution in frequency-domain
  - High-pass meta-filtering with window shape as transfer function
  - Amplified oszillations of neighbor Fourier coefficients
  - Phase difference $\longrightarrow \pi$.

Advantages

- Independent from neighbouring blocks
- 25% window size hop sufficient
- No phase unwrapping needed
- One FFT per frame sufficient

Disvantages

- Does not detect pitched non-percussive onsets

Thank you for your attention!