Onset Detection Revisited

Simon Dixon
simon.dixon@ofai.at

Austrian Research Institute for Artificial Intelligence
Vienna, Austria

9th International Conference on
Digital Audio Effects

O F A I

Outline

OFAI

Musical Onset Detection

- Aim: detect the start of musical tones
- Motivations: content-based analysis and retrieval of music information, automatic transcription, etc.
- This study:
  - verify and extend results of Bello et al. (TSAP 2005)
  - propose new onset detection functions
  - test on a large database of solo piano performances
  - test set 100 times larger than Bello et al.'s

Why is Onset Detection Difficult?

- Real-world data: complex polyphonic music
- Simultaneous or quasi-simultaneous notes
- Masking
- Chord asynchrony is at the limits of human perception (10's of milliseconds)
- How many onsets in a chord?
- Difficult to evaluate methods quantitatively

Previous Work

- Onset detection literature reviewed by Bello et al. (IEEE TSAP 2005)
- Empirical comparison of various methods using a standard data set
- Methods based on short-term spectral features
  - most widely used
  - most successful: winners of MIREX 2005 (see Downie, 2005)
- Further review by Collins (AES 118th Convention, 2005)

OFAI

Onset Detection Functions

- Peaks coincide with times of note onsets
- Low sampling rate w.r.t. audio
- Detect change in properties of audio signal
- Distinguish between various types of change:
  - onsets, offsets, vibrato, amplitude modulation, noise
- Basic ideas:
  - increase in energy in some frequency band(s)
  - irregularity in phase derivative in some frequency bands(s)
  - combinations of phase and amplitude/energy features

Audio Preprocessing

- Audio data: 44.1kHz mono
- STFT
- Hamming window
- Window length 2048 samples (46ms)
- Hop length 441 samples (10ms)
- Overlap 78.5%

Spectral Flux (SF): Existing Method

- Change in magnitude from frame to frame in each frequency bin
- Only positive changes
- Summed across frequency

Phase Deviation (PD): Existing Method

- Phase irregularities indicate transients
- Instantaneous frequency: first difference of phase
- Phase deviation: second difference of phase
- Averaged across frequency

Phase Deviation: Proposed Improvements

- Signal energy is concentrated in a few frequency bins
- Remaining components have low energy and random phase
- Proposal: weighted phase deviation (WPD)
  - phase deviation is multiplied by magnitude and summed over frequency
  - optionally, the WPD is normalised by dividing by the sum of magnitudes (NWPD)
- Joint consideration of magnitude and phase

Complex Domain (CD): Existing Method

- Joint consideration of magnitude and phase
- Search for departures from steady-state behaviour
- Steady-state: magnitude and instantaneous frequency remain constant
- Predict amplitude and phase based on 2 previous frames
- Sum absolute values of deviations (in complex plane) from predictions

Complex Domain: Proposed Improvements

- Complex domain method doesn't distinguish between increases and decreases in magnitude
- cf spectral flux, where only positive changes are considered
- Proposal: rectified complex domain function (RCD)
- Only sum differences where the magnitude is increasing

Methodology and Data

- Evaluation of multiple onsets (chords)
- Comparisons difficult (parameter settings for corresponding points on ROC curve)
- Precision (P), recall (R) and F-measure ($F = \frac{2PR}{P+R}$)
- Ground truth data is hard to find
- Computer-monitored piano (e.g. Bösendorfer SE) is an exception
- Alternative is hand-labelled data
- 2 data sets:
  - Hand-labelled short excerpts from various instruments (1060 onsets) from Bello et al. (2005)
  - 4 hours of complex piano music (106054 onsets)

Onset Selection

- Peak picking function is critical to onset detection performance
- Thresholds determine balance of false detections (false positives) and missed detections (false negatives)
- Optimal balance is application-specific
- Local maximum
- Peak higher than the local average (+ threshold)
- Peak not masked by previous higher peak (exponential decay)

OFAI

Results (F-measure)

	PN data	PP data	NP data	CM data	Piano
SF*	0.892	0.966	0.876	0.848	—
SF	0.952	0.984	0.967	0.882	0.964
PD*	0.957	0.976	0.871	0.776	—
PD	0.770	0.619	0.831	0.704	0.677
WPD	0.947	0.912	0.966	0.836	0.912
NWPD	0.938	0.971	0.958	0.879	0.944
CD	0.946	0.978	0.936	0.876	0.966
RCD	0.963	0.981	0.963	0.877	0.955

Discussion

- Large discrepancies with published results
  - implementation details affect results radically
  - PD* implemented with magnitude threshold
  - SF vs SF* shows improved peak-picking
- WPD and NWPD are significant improvements over PD
- RCD and CD differences are test-set specific
- Complex piano music has slightly worse results than PP data
- SF, CD, NWPD have similar overall performance
- Accuracy: SF 8.8ms; CD 12.8ms; NWPD 10.3ms
- SF is simplest and has most precise onset times