Onset Detection Revisited

Simon Dixon
simon.dixon@ofai.at

Austrian Research Institute for Artificial Intelligence
Vienna, Austria

9th International Conference on Digital Audio Effects
Outline

1 Background and Motivation
   - Introduction
   - Difficulties
   - Previous Work

2 Onset Detection Functions
   - Spectral Flux
   - Phase Deviation
   - Complex Domain

3 Evaluation and Results
   - Methodology and Data
   - Onset Selection
   - Results and Discussion
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)
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)
## Results (F-measure)

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