CCRMA MIR Workshop 2014
Beat–finding and Rhythm Analysis
Outline

- Modelling Rhythm Cognition.
- Onset-detection.
- Beat-tracking & Tempo-derivation.
  - Autocorrelation.
  - Beat Spectral approaches.
  - Histogram models.
- Meter determination.
- Applications, Exercises
Basic system overview

Segmentation
(Frames, Onsets, Beats, Bars, Chord Changes, etc)

Feature Extraction
(Time-based, spectral energy, MFCC, etc)

Analysis / Decision Making
(Classification, Clustering, etc)
Beat–finding and Tempo Derivation

- Why?
  - Tempo and Beat are strong discriminators in judgements of music similarity, and even genre (Tzanetakis & Cooke 2002, Dixon et. al 2004).
  - Understanding the beat facilitates understanding the importance other musical elements:
    - Relative importance of tonal features.
    - Diatonic or chromatic character of a piece.
    - Complexity of a piece.
  - Applications: musicology & ethnomusicology, automatic DJing, query by example, composition tools.
Example: Foot-tapping to singing

- Singing examples of Dutch folk songs from the "Onder de Groene Linde" collection (Meertens Institute).
- Uses continuous wavelet transform of rhythmic signals (Smith 1996, Smith & Honing 2008) to derive tactus:
  - Example 1: Original... + Accompaniment.
  - Example 2: ...Original + Accompaniment.
Modelling Rhythm

- “...the *systematic* patterning of sound in terms of timing, accent, and grouping.” (Patel 2008 p.96)

  - (Not always periodic patterns)

- Accent sources include: dynamics, melody, harmony, articulation, timbre, onset asynchrony etc.

- Consists of *hierarchical* and *figural* (proximal) temporal structures.
Measuring Beat

- Inter-Onset Intervals (IOI)
- Inter-Beat Interval (IBI)
- Tempo: frequency of the beat (BPM) = 1/IBI
Theory and Perception of Musical Time

- Multiple simultaneous levels of musical time
  - Tactus: the foot-tapping rate.
  - Tempo: estimated from tactus, typically median IBI.
  - Meter: Periodic perceived accentuation of beats.
  - Tatum: Shortest interval between events.

- Rubato – change in tempo during performance to emphasise structure.
Rhythmic Strata

- Musical rhythm can be considered as composed of a hierarchy of temporal levels or strata (Yeston 1976, Lerdahl & Jackendoff 1983, Clarke 1987, Jones & Boltz 1989).

From Jones & Boltz ‘89
Metrical Structure

- Meter is expressed as a hierarchical grouping in time. E.g.
  Subdivision of 4/4 (4 beats to the bar):

![Salience Hierarchy for (2 2 2 2) meter](image)
• Meter is expressed in Western music as time–signatures (4/4, 3/4 etc).

Subdivision of 4/4 (4 beats to the bar):
Meter

- Subdivision of 3/4 (3 beats to the bar):

![Salience Hierarchy for (3 2 2) meter](image-url)
Meter

- Subdivision of 6/8:

![Salience Hierarchy for (2 3 2) meter](image)
Hierarchical Grouping: Meter

- Meters are argued to arise from the interaction between temporal levels (Yeston 1976).
  - Therefore a meter implies two frequencies: the pulse rate and the measure (“bar”) rate.
- The tactus is considered as the most salient hierarchical level, consistent with the notated meter, or the foot tapping rate (Desain & Honing 1994).
Mental schemas for Meter

- **Metrical Profiles** (Palmer & Krumhansl 1990)
  - Pre-established mental frameworks ("schemas") for musical meter are used during listening.

From Palmer & Krumhansl (1990). Mean goodness-of-fit ratings for musicians (solid line) and nonmusicians (dashed line).
Syncopation

• “...some kind of conflict between [phenomenal] accents and meter” (Temperely 2001, p.239).

Syncopation

- Listener judgements of musical complexity are correlated with degree of syncopation (i.e. note location within the beat) (Shmulevich & Povel 2000, Smith & Honing 2006).

- Compared judgements against formal model of syncopation (Longuet-Higgins & Lee 1984).
Active Rhythm Perception

- Viewed as a resonance between top down and bottom-up processes (see e.g Desain & Honing 2001):
Onset-detection vs. Beat-detection

• Traditionally beat detection relied on accurate onset detection.
  - i.e from MIDI data for **Score Following** (Dannenberg 1991, Cont 2009).

• This can be difficult for MIR from polyphonic audio recordings.
  - A higher freq. **Onset Detection Function** from the entire audio signal can be used for beat tracking without all onsets being detected (Schloss 1985, Goto & Muraoka 1994, Scheirer 1998).
The Onset Detection Function

- Represents:
  - Ideal: Each note that contributes to the beat.
  - Practice: Combined envelopes of all notes.

- Tends to emphasise:
  - strong transients (i.e. impulsive sounds)
  - loud notes
  - bass notes
  - wide-band spectrum events (e.g. snare drums).
Dixon’s Envelope Onset Detection
Example Onset Detection

- Pre-processing
- Filtering
- Down-sampling
- Difference function
Common ODF methods

- e.g (Bello et. al 2005, Dixon 2007, Peeters 2007)
  • Optional pre-rectification filtering.
  • Envelope mixture from rectification/energy.
  • Smoothing of envelope (LP filter).
  • Down-sampling for data reduction.
  • \( \frac{d(\log E)}{dt} \) highlights perceived impulses.
  • Weighting higher frequencies captures wide-band events.
  • Spectral difference between STFT frames.
Existing Beat tracking Models

- Parsing metrical grammars (Longuet-Higgins and Lee 1982).
- Forward projection of likelihood (Desain 1992).
Approaches to beat tracking considered

- Autocorrelation
  - Finding Periodicity in the ODF.

- Beat Spectrum approaches:
  - Spectrum of the ODF.
  - Multi-resolution representation of ODF.

- Dynamic Programming approaches.
  - Efficient selection of correct beat interval.
Autocorrelation of ODF

- AC peaks ⇒ time lags where signal is most similar to itself.
- Captures periodicities of ODF.
- Does not capture rubato well.
- OK for metronomic music, not for those with variation in tempo.
Windowed RMS and its Autocorrelation (for drum loop)

Max peak = 2-bar loop

1st peak = 1/8 note

1/4 note

Max peak = 2-bar loop
Beat spectrum methods (Scheirer 1998)

- Filterbanks of tuned resonators (i.e. “rhythmic reverb”) of the ODF.
- Resonator whose resonant $F$ matches rate of ODF modulation will phase-lock.
- Resonator outputs of common freq summed across subbands:

$$T = \arg \max_r \sum_s F_{rs}$$
Beat Tracking by Peeters (2007)

Figure 1: Flowchart of our system for tempo, meter estimation, and beat marking.
• Filtered, rectified spectral energy envelope
  – Onset detection function.
• Combined Fourier & autocorrelation analysis
  – DFT of ODF, ACF of ODF
  – ACF result mapped into Fourier domain.
  – DFT * Freq(ACF) – disambiguates periodicities.
  – Octave errors occur in two different domains.
Viterbi decoding of joint estimates of meter and tempo.

Peeters 2007
Dynamic Programming (Ellis 2007)

- Goal to generate beat times that match onsets and have near constant IBI.

\[
C(\{t_i\}) = \sum_{i=1}^{N} O(t_i) + \alpha \sum_{i=2}^{N} F(t_i - t_{i-1}, \tau_p)
\]

- \(F(\Delta t, \tau) = -\log(\text{actual IBI/ideal IBI})^2\).

- Ideal IBI from tempo estimation from weighted autocorrelation.

- Recursively calculates max \(C^*(t)\) starting from \(t=0-2\tau\), finding times of max(\(F + C^*(\tau)\)).

- Chooses final max \(C^*(t)\) from last interval, backtraces the saved times.
Beat Histograms

- Summarises rhythmic behaviour of a piece for similarity measures, classification etc.
- Pampalk, Dixon & Widmer (2003)
  - Uses summation of comb filters of Scheirer, not just argmax, for comparison.
- Tempo histogram is weighted using a preference model (van Noorden & Moelants 1999).
- PCA used to reduce 2000 → 60 dimensions for matching.

(from Scheirer 1998)
Beat Histograms (Tzanetakis and Cook, 2002)

- Similar approach using Autocorrelation.
- Add the amplitudes of the top 3 AC peaks to histogram at each frame.
- Beat histograms are reducible to single features including sum and peak/mean.
Fluctuation Patterns

- Also summarises rhythmic behaviour.
- FFT of envelope: the fluctuation (AM) frequency of the perceived loudness of critical bands (log spectral) (represented on the Bark scale).
- 20 Bark x 60 BF matrix $\Rightarrow$ PCA for matching

Median of the fluctuation patterns of examples of (L-R) Heavy Metal, Dance and Pop. Y axis shows critical bands (Bark 1-20), X axis shows beat frequencies 0-10Hz (0-600BPM) From Pampalk, Rauber & Merkl, (2002)
Meter estimation

- Requires measure (“bar”) period and phase (downbeat) identification.
- Measure period reasonably successful, albeit with octave errors.
- Downbeat identification much harder!
- Genre dependent.
Table 1. Characteristics of some meter estimation systems.

<table>
<thead>
<tr>
<th>Reference</th>
<th>Input</th>
<th>Output</th>
<th>Technique</th>
</tr>
</thead>
<tbody>
<tr>
<td>Temperley &amp; Sleator (1999)</td>
<td>MIDI</td>
<td>Meter, time quantization</td>
<td>Rule-based approach; implementation of the preference rules in (Lerdahl et al., 1983)</td>
</tr>
<tr>
<td>Dixon (2001)</td>
<td>MIDI, audio</td>
<td>Tactus</td>
<td>First find periods using IOI histogram, then phases using multiple agents</td>
</tr>
<tr>
<td>Raphael (2001)</td>
<td>MIDI, audio</td>
<td>Tactus, time quantization</td>
<td>Probabilistic generative model for onset times; MAP estimation (Viterbi)</td>
</tr>
<tr>
<td>Cemgil &amp; Kappen (2003)</td>
<td>MIDI</td>
<td>Tactus, time quantization</td>
<td>Probabilistic generative model for onset times; sequential Monte Carlo methods</td>
</tr>
<tr>
<td>Goto &amp; Mur-aoka (1995, 1997)</td>
<td>Audio</td>
<td>Meter</td>
<td>Extract onset components; IOI histogram; multiple tracking agents</td>
</tr>
<tr>
<td>Scheirer (1998)</td>
<td>Audio</td>
<td>Tactus</td>
<td>Bank of comb filters to analyze periodicity of power envelopes at six subbands</td>
</tr>
<tr>
<td>Laroche (2001)</td>
<td>Audio</td>
<td>Tactus, swing</td>
<td>Extract discrete onsets; maximum-likelihood estimation</td>
</tr>
<tr>
<td>Sethares &amp; Staley (2001)</td>
<td>Audio</td>
<td>Meter</td>
<td>Calculate RMS-energies at 1/3-octave subbands; apply a periodicity transform</td>
</tr>
<tr>
<td>Gouyon et al. (2002)</td>
<td>Audio</td>
<td>Tatum</td>
<td>First find periods (IOI histogram), then phases by matching isochronous pattern</td>
</tr>
<tr>
<td>Klapuri et al. (to appear)</td>
<td>Audio</td>
<td>Meter</td>
<td>Measure degree of accentuation; bank of comb filters; probabilistic model</td>
</tr>
</tbody>
</table>
Joint estimation of chord change and downbeat (Papadopoulos & Peeters 2008)

- **Hidden Markov Model:**
  - States: 24 Major & Minor triads * 4 positions within the Measure (pim) for (4/4 time signature).
  - Computes chroma features at each beat.
  - Assumes independence between beat position and chord type: $P(O|s) = P(O|c) P(O|pim)$
  - Transition probabilities enforce sequential beats & likelihood of chord transitions.

- **Optimal state determined by Viterbi decoding.**
  - Chord progression detection improved using metrical knowledge.
  - Identification of downbeats aided by harmonic information.
Review

• Modeling rhythm requires representing perception.
• Onset detection functions capture significant events.
• Multiple approaches to beat-tracking represent competing perceptual models.
• Beat-tracking enables higher-level rhythmic features (FP, BH).
• Beat-tracking enables multi-modal estimation (e.g., down-beat from chords).