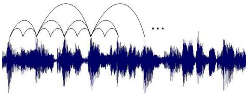
ISMIR 2006 TUTORIAL: Computational Rhythm Description

Fabien Gouyon   Simon Dixon

Austrian Research Institute for Artificial Intelligence, Vienna
http://www.ofai.at/~fabien.gouyon
http://www.ofai.at/~simon.dixon

7th International Conference on Music Information Retrieval



Outline of the Tutorial

- Introductory Concepts: Rhythm, Meter, Tempo and Timing
- Functional Framework

Coffee Break

- Evaluation of Rhythm Description Systems
- MIR Applications of Rhythm Description
- Some ideas

Part I

Introductory Concepts: Rhythm, Meter, Tempo and Timing

Outline

- Introduction
- Rhythm
  - Meter
  - Tempo
  - Timing

O F A I

The Big Picture

- Music = Organised Sound
- Traditional analysis looks at 4 main components of music:
  - melody
  - rhythm
  - harmony
  - timbre

OFAI

Music Representation

- Score
  - Discrete
  - High level of abstraction (e.g. timing not specified)
  - Structure is explicit (bars, phrases)
  - Not suitable for detailed performance information
- MIDI
  - Discrete
  - Medium level of abstraction
  - Timing is explicit, structure can be partly specified
  - Suitable for keyboard performance representation
- Audio
  - Continuous (for our purposes)
  - Low level of abstraction
  - Timing and structure are implicit

OFAI

Event-Based Representation of Music

- Simple and efficient
- e.g. MIDI
  - Events are durationless (i.e. occur at a point in time)
  - Musical notes consist of a start event (onset or *note-on* event) and an end event (offset, *note-off* event)
  - Notes have scalar attributes
    e.g. for pitch, dynamics (velocity)
  - Difficult to represent intra-note expression
    e.g. vibrato, dynamics
- Extracting an event representation from an audio file is difficult
  - e.g. onset detection, melody extraction, transcription

What is Rhythm?

- Music is a temporal phenomenon
- Rhythm refers to medium and large-scale temporal phenomena
    - i.e. at the event level
- Rhythm has the follow components:
    - Timing: *when* events occur
    - Tempo: *how often* events occur
    - Meter: *what structure* best describes the event occurrences
    - Grouping: phrase structure (not discussed)
- References: Cooper and Meyer (1960); Lerdahl and Jackendoff (1983); Honing (2001)

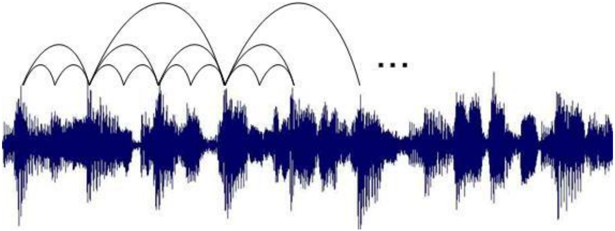
Meter: Beat and Pulse

- *Pulse*: regularly spaced sequence of accents
  - can also refer to an element of such a sequence
  - *beat* and *pulse* are often used interchangeably, but ...
  - *pulse* $\longrightarrow$ a sequence
  - *beat* $\longrightarrow$ an element
- Explicit in score (time signature, bar lines)
- Implicit in audio
- Multiple pulses can exist simultaneously

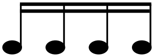
Metrical Structure

- Hierarchical set of pulses
- Each pulse defines a *metrical level*
- Higher metrical levels correspond to longer time divisions
- Well-formedness rules (Lerdahl and Jackendoff, 1983)
  - The beats at each metrical level are equally spaced
  - There is a beat at some metrical level for every musical note
  - Each beat at one metrical level is an element of the pulses at all lower metrical levels
  - A beat at one metrical level which is also a beat at the next highest level is called a *downbeat*; other beats are called *upbeats*
- Different from grouping (phrase) structure
- Doesn't describe performed music

Metrical Structure



Meter: Notation

- all notes are fractions of an arbitrary duration
- whole note: 
- half note: 
- quarter note: 
- eighth notes: 
- sixteenth notes: 
- a dot after the note adds 50% to the duration
- a curve joining two note symbols sums their duration

Notation: Time Signature

- The time signature describes part of the metrical structure
- It consists of 2 integers arranged vertically, e.g. $\frac{4}{4}$ or $\frac{6}{8}$
  - these determine the relationships between metrical levels
  - the lower number is the units of the *nominal beat level* (e.g. 4 for a quarter note)
  - the upper number is the count of how many units per bar (measure)
  - compound time: if the upper number is divisible by 3, an intermediate metrical level is implied (grouping the nominal beats in 3's)
- It is specified in the score, but can't be determined unambiguously from audio

Tempo

- Tempo is the rate of a pulse (e.g. the nominal beat level)
- Usually expressed in beats per minute (BPM), but the inter-beat interval (IBI) can also be used (e.g. milliseconds per beat)
- Problems with measuring tempo:
  - Variations in tempo
  - Choice of metrical level
  - Tempo is a perceptual value (strictly speaking), so it can only be determined empirically (cf pitch)

Tempo Variations

- Humans do not play at a constant rate
- Instantaneous tempo doesn't really exist
- Tempo can at best be expressed as a central tendency
  - Basic tempo: mean, mode (Repp, 1994)
  - Local tempo: calculated with moving window
  - Instantaneous tempo: limit as window size approaches 0
- Not all deviations from metrical timing are tempo changes

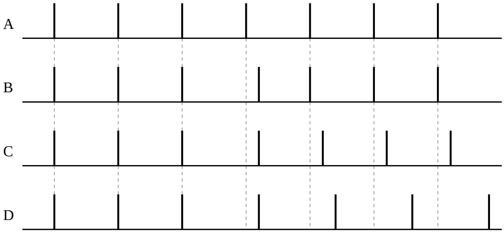
Tempo: Choice of Metrical Level

- Tapping experiments
  - people prefer moderate tempos (Parncutt, 1994; van Noorden and Moelants, 1999)
  - people tap at different metrical levels
  - results are not restricted to tapping (Dixon et al., 2006)
- The nominal beat level (defined by the time signature) might not correspond to the perceptual tempo
  - but it might be the best approximation we have
- Affected by factors such as note density, musical training

OFAI

Timing

- Not all deviations from metrical timing are tempo changes



- Nominally on-the-beat notes don't occur *on* the beat
  - difference between notation and perception
  - "groove", "on top of the beat", "behind the beat", etc.
  - systematic deviations (e.g. swing)
  - expressive timing

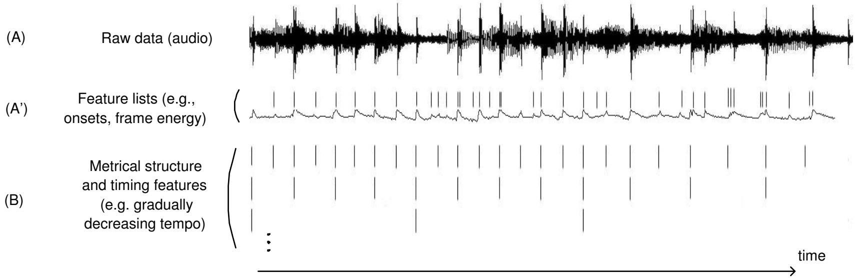
Problems with Representation of Performance Timing

- Most representations and approaches ignore performance timing
- Mathematically underspecified — too many degrees of freedom
- e.g. Tempo curve (Desain and Honing, 1991a)
- Causal analysis is not possible
- References: Desain and Honing (1991b); Honing (2001); Dixon et al. (2006)

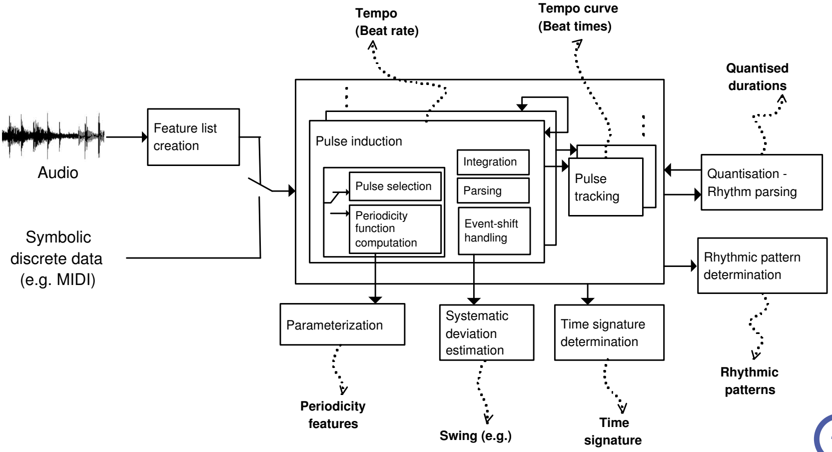
Part II

Functional Framework

Automatic Rhythm Description



Functional Units of Rhythm Description Framework



Extension of (Gouyon and Dixon, 2005b)

Outline

- Input Data
- Rhythm periodicity functions
- Pulse induction
- Beat Tracking
- Extraction of Higher Level Rhythmic Features

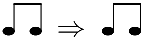
Input Data

- Different type of inputs:
  - discrete data, e.g.:
    - parsed score (Longuet-Higgins and Lee, 1982; Brown, 1993)
    - MIDI data (Cemgil et al., 2000a)
  - continuous audio data (Schloss, 1985)
- First step: Parsing data into a feature list conveying (hopefully) most relevant information to rhythmic analysis

Event-wise features

- Onset time (Longuet-Higgins and Lee, 1982; Desain and Honing, 1989)
- Duration (Brown, 1993; Parncutt, 1994)
- Relative amplitude (Smith, 1996; Meudic, 2002)
- Pitch (Chowning et al., 1984; Dixon and Cambouropoulos, 2000)
- Chords (Rosenthal, 1992b)
- Percussive instrument classes (Goto and Muraoka, 1995; Gouyon, 2000)

Event-wise features

- When processing continuous audio data
  $\Rightarrow$ Transcription audio-to-MIDI (Klapuri, 2004; Bello, 2003)
- Onset detection literature (Klapuri, 1999; Dixon, 2006)
  
- Pitch and chord estimation (Gómez, 2006)
- Monophonic audio data
  - $\longrightarrow$ Monophonic MIDI file
- Polyphonic audio data
  - $\longrightarrow$ Stream segregation and transcription
  - $\longrightarrow$ "Summary events"
- Very challenging task

Frame-wise features

- Lower level of abstraction might be more relevant perceptually (Honing, 1993), criticism of the “transcriptive metaphor” (Scheirer, 2000)
- Frame size = 10-20 ms, hop size = 0-50%
  - energy, energy in low freq. band (low drum, bass) (Wold et al., 1999; Alghoniemy and Tewfik, 1999)
  - energy in different freq. bands (Sethares and Staley, 2001; Dixon et al., 2003)
  - energy variations in freq. bands (Scheirer, 1998)
  - spectral flux (Foote and Uchihashi, 2001; Laroche, 2003)
  - reassigned spectral flux (Peeters, in press)
  - onset detection features (Davies and Plumbley, 2005)
  - spectral features (Sethares et al., 2005; Gouyon et al., in press)

O F A I

Frame-wise features

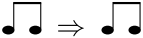
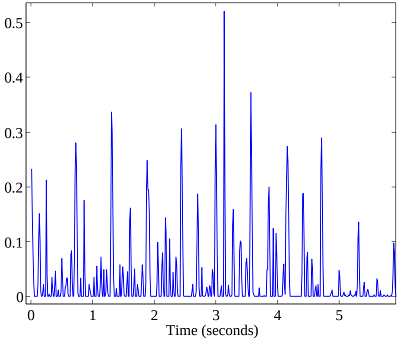


Figure: Normalised energy variation in low-pass filter



Beat-wise features

- Compute features over the time-span defined by 2 consecutive beats.
- Requires knowledge of a lower metrical level, e.g. Tatum for Beat, Beat for Measure.
    - chord changes at the 1/4 note level (Goto and Muraoka, 1999)
    - spectral features at the Tatum level (Seppänen, 2001a; Gouyon and Herrera, 2003a; Uhle et al., 2004)
    - temporal features, e.g. IBI temporal centroid (Gouyon and Herrera, 2003b)

Rhythm periodicity functions

- Representation of periodicities in feature list(s)
- Continuous function representing magnitude –or salience (Parncutt, 1994)– vs. period –or frequency–
- Diverse pre- and post-processing:
  - scaling with tempo preference distribution (Parncutt, 1994; Todd et al., 2002; Moelants, 2002)
  - encoding aspects of metrical hierarchy (e.g. influence of some periodicities on others)
    - favoring rationally-related periodicities
    - seeking periodicities in Periodicity Function
  - emphasising most recent samples
    - use of a window (Desain and de Vos, 1990)
    - intrinsic behavior of comb filter, Tempogram

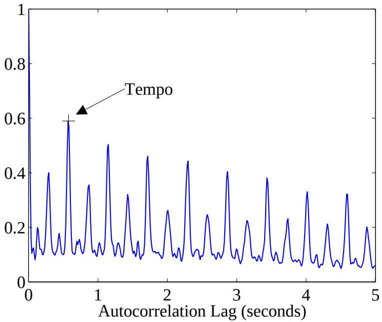
Examples: Autocorrelation

Most commonly used, e.g. Desain and de Vos (1990); Brown (1993); Scheirer (1997); Dixon et al. (2003)
Measures feature list self-similarity vs time lag

$$r(\tau) = \sum_{n=0}^{N-\tau-1} x(n)x(n+\tau) \qquad \forall \tau \in \{0 \cdots U\}$$

$x(n)$: feature list, N: number of samples
τ: lag
U: upper limit
$N - \tau$: integration time
Normalisation $\Rightarrow r(0) = 1$

Examples: Autocorrelation



(Feature: normalised energy variation in low-pass filter)

Examples: Autocorrelation

Variants:

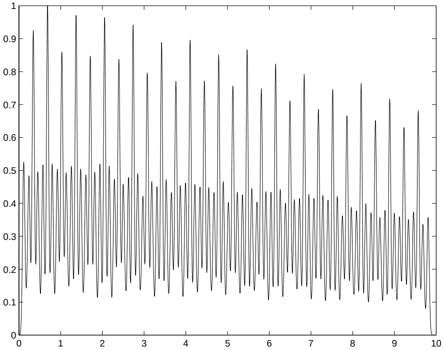
- Autocorrelation Phase Matrix (Eck, in press)
- Narrowed ACF (Brown and Puckette, 1989)
- "Phase-Preserving" Narrowed ACF (Vercoe, 1997)
- Sum or correlation over similarity matrix (Foote and Uchihashi, 2001)

Examples: Time interval histogram

Seppänen (2001b); Gouyon et al. (2002)

- Compute onsets
- Compute IOIs
- Build IOI histogram
- Smoothing with e.g. Gaussian window
- See IOI clustering scheme by Dixon (2001a)

Examples: Time interval histogram



(Feature: Onset time+Dynamics)

Examples: Pulse Matching

Gouyon et al. (2002)

- With onset list
  - generate pulse grids (enumerating a set of possible pulse periods and phases)
  - compute two error functions, e.g. Two-Way Mismatch error (Maher and Beauchamp, 1993)
    1. how well do onsets explain pulses? (Positive evidence)
    2. how well do pulses explain onsets? (Negative evidence)
  - linear combination
  - seek global minimum
- With continuous feature list
  - compute inner product (Laroche, 2003)
  - comparable to Tempogram (Cemgil et al., 2001)

Examples: Others

- Comb filterbank (Scheirer, 1998; Klapuri et al., 2006)
- Fourier transform (Blum et al., 1999)
- Combined Fourier transform and Autocorrelation (Peeters, in press)
- Wavelets (Smith, 1996)
- Periodicity transform (Sethares and Staley, 2001)
- Tempogram (Cemgil et al., 2001)
- Beat histogram (Tzanetakis and Cook, 2002; Pampalk et al., 2003)
- Fluctuation patterns (Pampalk et al., 2002; Pampalk, 2006; Lidy and Rauber, 2005)

“Best” periodicity function?

- Is there a best way to emphasise periodicities?
- Does it depend on the input feature?
- Does it depend on the purpose?

Periodicity features

Low-level descriptors of rhythm periodicity functions

- Whole function (Foote et al., 2002)
- Sum (Tzanetakis and Cook, 2002; Pampalk, 2006)
- Peak positions (Dixon et al., 2003; Tzanetakis and Cook, 2002)
- Peak amplitudes, ratios (Tzanetakis and Cook, 2002; Gouyon et al., 2004)
- Selected statistics (higher-order moments, flatness, centroid, etc.) (Gouyon et al., 2004; Pampalk, 2006)

Periodicity features

Applications:

- Genre classification
- Rhythm similarity
- Speech/Music Discrimination (Scheirer and Slaney, 1997)
- etc.

Pulse induction

- Select a pulse period, e.g. tempo, tatum $\Rightarrow$ 1 number
- Provide input to beat tracker (Desain and Honing, 1999)
- Assumption: pulse period and phase are stable
  - on the whole data (tempo almost constant all over, suitable to off-line applications)
  - on part of the data (e.g. 5 s, suitable for streaming applications)

Rhythm periodicity function processing

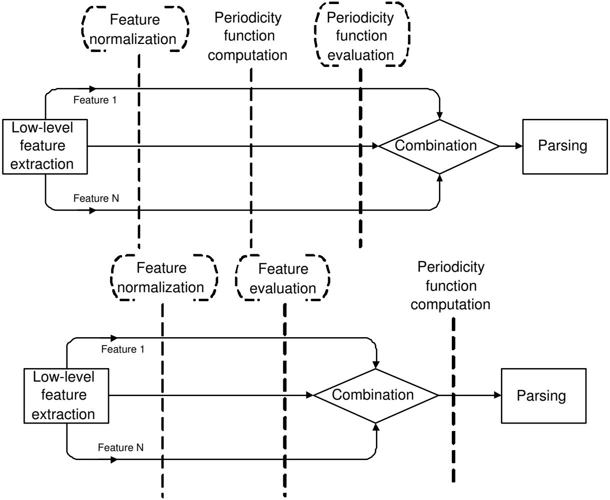
- Handling short-time deviations
- Combining multiple information sources
- Parsing

Handling short-time deviations

- Feature periodicities are always approximate
- Problem especially with discrete data (e.g. onset lists)
  - smooth out deviations, consider "tolerance interval"
    - rectangular window (Longuet-Higgins, 1987; Dixon, 2001a)
    - Gaussian window (Schloss, 1985)
    - window length may depend on IOI (Dixon et al., 2003; Chung, 1989)
  - handle deviations to derive systematic patterns
    - swing

O F A I

Combining multiple information sources



Combining multiple information sources

- If multiple features are used (e.g. energy in diverse freq. bands)
  - first compute rhythm periodicity functions (RPFs), then combine
  - first combine, then compute RPF
- Evaluate worth of each feature
  e.g. periodic $\Leftrightarrow$ good
  - evaluate "peakiness" of RPFs
  - evaluate variance of RPFs
  - evaluate periodicity of RPFs
- Normalize features
- "Combination"
  - (weighted) sum or product
  - considered jointly with Parsing...

Parsing

- Continuous RPF
  $\Rightarrow$ Pulse period, 1 number
- Max peak: Tactus (Schloss, 1985)
- Max peak in one-octave region, e.g. 61-120 BPM
- Peak $>$ all previous peaks & all subsequent peaks up to twice its period (Brown, 1993)
- Consider constraints posed by metrical hierarchy
  - consider only periodic peaks (Gouyon and Herrera, 2003a)
  - collect peaks from several RPFs, score all Tactus/Measure hypotheses (Dixon et al., 2003)
  - beat track several salient peaks, keep most regular track (Dixon, 2001a)
  - probabilistic framework (Klapuri et al., 2006)

Parsing - Future Work

- Difficulty to compute, but also to define the “right” pulse $\Rightarrow$ Problem for evaluations when no reference score is available
- Design rhythm periodicity function whose peak amplitude would correspond to perceptual salience (McKinney and Moelants, 2004)
- New algorithms for combining and parsing features or periodicity functions

Pulse selection

- Evaluating the salience of a restricted number of periodicities
- Suitable only to discrete data
- Instance-based approach
  - first two events (Longuet-Higgins and Lee, 1982)
  - first two agreeing IOIs (Dannenberg and Mont-Reynaud, 1987)
- Pulse-matching
  - positive evidence: number events that coincide with beats
  - negative evidence: number of beats with no corresponding event
- Usually not efficient, difficulty translated to subsequent tracking process

Beat Tracking

- Complementary process to tempo induction
- Fit a grid to the events (resp. features)
  - basic assumption: co-occurence of events and beats
  - e.g. by correlation with a pulse train
- Constant tempo and metrical timing are not assumed
  - grid must be flexible
  - short term deviations from periodicity
  - moderate changes in tempo
- Reconciliation of predictions and observations
- Balance:
  - reactiveness (responsiveness to change)
  - inertia (stability, importance attached to past context)

Beat Tracking Approaches

- Top down and bottom up approaches
- On-line and off-line approaches
- High-level (style-specific) knowledge vs generality
- Rule-based (Longuet-Higgins and Lee, 1982, 1984; Lerdahl and Jackendoff, 1983; Desain and Honing, 1999)
- Oscillators (Povel and Essens, 1985; Large and Kolen, 1994; McAuley, 1995; Gasser et al., 1999; Eck, 2000)
- Multiple hypotheses / agents (Allen and Dannenberg, 1990; Rosenthal, 1992a; Rowe, 1992; Goto and Muraoka, 1995, 1999; Dixon, 2001a)
- Filter-bank (Scheirer, 1998)
- Repeated induction (Chung, 1989; Scheirer, 1998)
- Dynamical systems (Cemgil and Kappen, 2001)

State Model Framework for Beat Tracking

- set of state variables
- initial situation (initial values of variables)
- observations (data)
- goal situation (the best explanation for the observations)
- set of actions (adapting the state variables to reach the goal situation)
- methods to evaluate actions

State Model: State Variables

- pulse period (tempo)
- pulse phase (beat times)
  - expressed as time of first beat (constant tempo) or current beat (variable tempo)
- current metrical position (models of complete metrical structure)
- confidence measure (multiple hypothesis models)

State Model: Observations

- All events or events near predicted beats
- Onset times, durations, inter-onset intervals (IOIs)
  - equivalent only for monophonic data without rests
  - longer notes are more indicative of beats than shorter notes
- Dynamics
  - louder notes are more indicative of beats than quieter notes
  - difficult to measure (combination/separation)
- Pitch and other features
  - lower notes are more indicative of beats than higher notes
  - particular instruments are good indicators of beats (e.g. snare drum)
  - harmonic change can indicate a high level metrical boundary

State Models: Actions and Evaluation

A simple beat tracker:

- Predict the next beat location based on current beat and beat period
- Choose closest event and update state variables accordingly
- Evaluate actions on the basis of agreement with prediction

OFAI

Example 1: Rule-based Approach

- Longuet-Higgins and Lee (1982)
- Meter is regarded as a generative grammar
  - A rhythmic pattern is a parse tree
- Parsing rules, based on musical intuitions:
  - CONFLATE: when an expectation is fulfilled, find a higher metrical level by doubling the period
  - STRETCH: when a note is found that is longer than the note on the last beat, increase the beat period so that the longer note is on the beat
  - UPDATE: when a long note occurs near the beginning, adjust the phase so that the long note occurs on the beat
  - LONGNOTE: when a note is longer than the beat period, update the beat period to the duration of the note
  - An upper limit is placed on the beat period
- Biased towards reactiveness

Example 2: Metrical Parsing

- Dannenberg and Mont-Reynaud (1987)
- On-line algorithm
- All incoming events are assigned to a metrical position
- Deviations serve to update period
- Update weight determined by position in metrical structure
- Reactiveness/inertia adjusted with *decay* parameter
- Extended to track multiple hypotheses (Allen and Dannenberg, 1990)
  - delay commitment to a particular metrical interpretation
  - greater robustness against errors
  - less reactive
- Evaluate each hypothesis (credibility)
- Heuristic pruning based on musical knowledge
- Dynamic programming (Temperley and Sleator, 1999)

Example 3: Coupled Oscillators

- Large and Kolen (1994)
- Entrainment: the period and phase of the *driven* oscillator are adjusted according to the *driving* signal (a pattern of onsets) so that the oscillator synchronises with its beat
- Oscillators are only affected at certain points in their cycle (near expected beats)
- Multiple oscillators entrain simultaneously
- Adaptation of period and phase depends on coupling strength (determines reactiveness/inertia balance)
- Networks of connected oscillators could model metrical structure

Example 4: Multiple Agents

- Goto and Muraoka (1995)
- Real-time beat tracking of audio signals
- Finds beats at quarter and half note levels
- Detects onsets, specifically labelling bass and snare drums
- Matches drum patterns with templates to avoid doubling errors and phase errors
- 14 pairs of agents receive different onset information
- Beat times are predicted using auto-correlation (tempo) and cross-correlation (phase)
- Agents evaluate their reliability based on fulfilment of predictions
- Limited to pop music with drums, $\frac{4}{4}$ time, 65–185 BPM, almost constant tempo

Example 5: Comb Filterbank

- Scheirer (1998)
- Causal analysis
- Audio is split into 6 octave-wide frequency bands, low-pass filtered, differentiated and half-wave rectified
- Each band is passed through a comb filterbank (150 filters from 60–180 BPM)
- Filter outputs are summed across bands
- Maximum filter output determines tempo
- Filter states are examined to determine phase (beat times)
- Problem with continuity when tempo changes
- Tempo evolution determined by change of maximal filter
- Multiple hypotheses: best path (Laroche, 2003)

Time Signature Determination

- Parsing the periodicity function
  - two largest peaks are the bar and beat levels (Brown, 1993)
  - evaluate all pairs of peaks as bar/beat hypotheses (Dixon et al., 2003)
- Parsing all events into a metrical structure (Temperley and Sleator, 1999)
- Obtain metrical levels separately (Gouyon and Herrera, 2003b)
- Using style-specific features
  - chord changes as bar indicators (Goto and Muraoka, 1999)
- Probabilistic model (Klapuri et al., 2006)

Rhythm Parsing and Quantisation

- Assign a position in the metrical structure for every note
- Important for notation (transcription)
- By-product of generating complete metrical hierarchy
- Discard timing of notes (ahead of / behind the beat)
- Should model musical context (e.g. triplets, tempo changes) (Cemgil et al., 2000b)
- Simultaneous tracking and parsing has advantages
- e.g. Probabilistic models (Raphael, 2002; Cemgil and Kappen, 2003)

Systematic Deviations

- Studies of musical performance reveal systematic deviations from metrical timing
- Implicit understanding concerning interpretation of notation
- e.g. swing: alternating long-short pattern in jazz (usually at 8th note level)
- Periodicity functions give distribution but not order
- Joint estimation of tempo, phase and swing (Laroche, 2001)

Rhythm Patterns

- Distribution of time intervals (ignoring order):
  - beat histogram (Tzanetakis and Cook, 2002)
  - modulation energy (McKinney and Breebaart, 2003)
  - periodicity distribution (Dixon et al., 2003)
- Temporal order defines patterns (musically important!)
- Query by tapping (Chen and Chen, 1998)
  - MIDI data
  - identity
- Comparison of patterns (Paulus and Klapuri, 2002)
  - patterns extracted from audio data
  - similarity of patterns measured by dynamic time warping
- Characterisation and classification by rhythm patterns (Dixon et al., 2004)

Coffee Break

Part III

Evaluation of Rhythm Description Systems

- Model improvements on the long term are bounded to systematic evaluations (see e.g. in text retrieval, speech recognition, machine learning, video retrieval)
- Often through contests, benchmarks
- Little attention in Music Technology
- Acknowledgment in MIR community (Downie, 2002)
- In the rhythm field:
  - tempo induction
  - beat tracking

Outline

- Methodology
  - Annotations
  - Data
  - Metrics
- ISMIR 2004 Audio Description Contest
  - Audio Tempo Induction
  - Rhythm Classification
- MIREX
  - MIREX 2005
  - MIREX 2006
- The Future
  - More Benchmarks
  - Better Benchmarks

Methodology

- Systematic evaluations of competing models are desirable
  They require:
  - an agreement on the manner of representing and annotating relevant information about data
  - reference examples of correct analyses, that is, large and publicly available annotated data sets
  - agreed evaluation metrics
  - (infrastructure)
- Efforts still needed on of all these points

Annotations

- Tempo in BPM
- Beats
- Meter
- Annotation tools:
  - Enhanced Wavesurfer (manual)
  - BeatRoot (semi-automatic)
  - QMUL's Sonic Visualizer (semi-automatic)
  - Other free or commercial audio or MIDI editors (manual)
- Several periodicities with respective saliences
- Perceptual tempo categories ("slow", "fast", "very fast", etc.)
- Complete score

Annotated Data - MIDI

- MIDI performances of Beatles songs (Cemgil et al., 2001), `http://www.nici.kun.nl/mmm/archives/`: Score-matched MIDI, ˜200 performances of 2 Beatles songs by 12 pianists, several tempo conditions
- “Kostka-Payne” corpus (Temperley, 2004), `ftp://ftp.cs.cmu.edu/usr/ftp/usr/sleator/melisma2003`: Score-matched MIDI, 46 pieces with metronomical timing and 16 performed pieces, “common-practice” repertoire music

Annotated Data - Audio

- RWC Popular Music Database `http://staff.aist.go.jp/m.goto/RWC-MDB/`: Audio, 100 items, tempo ("rough estimates")
- ISMIR 2004 data (Gouyon et al., 2006), `http://www.ismir2004.ismir.net/ISMIR_Contest.html`: Audio, $>$ 1000 items (+links to $>$ 2000), tempo
- MIREX 2005-2006 training data `http://www.music-ir.org/evaluation/MIREX/data/2006/beat/`: Audio, 20 items, 2 tempi + relative salience, beats

Evaluation Metrics

- Multidimensional, depends on
  - dimension under study, e.g.
    - tempo
    - beats
    - several metrical levels
    - quantised durations
  - criteria, e.g.
    - time precision (e.g. for performance research)
    - robustness
    - metrical level precision and stability
    - computational efficiency
    - latency
    - perceptual or cognitive validity
  - richness (and accuracy) of annotations
    - depend partly on input data type
    - hand-labelling effort (and care)
    - what level of resolution is meaningful?

Evaluation Metrics

- Comparison annotated and computed beats (Goto and Muraoka, 1997; Dixon, 2001b; Cemgil et al., 2001; Klapuri et al., 2006)
  - cumulated distances in beat pairs, false-positives, missed
  - longest correctly tracked period
  - particular treatment to metrical level errors (e.g. $2\times$)
- Matching notes/metrical levels (Temperley, 2004)
  - requires great annotation effort (complete transcriptions)
  - unrealistic for audio signals (manual & automatic)
- Statistical significance

ISMIR 2004 Audio Description Contest

- First large-scale comparison of algorithms
  - Genre Classification/Artist Identification
  - Melody Extraction
  - Tempo Induction
  - Rhythm Classification
- Cano et al. (2006), `http://ismir2004.ismir.net/ISMIR_Contest.html`

Audio Tempo Induction - Outline

- Compare state-of-the-art algorithms in the task of inducing the basic tempo (i.e. a scalar, in BPM) from audio signals
- 12 algorithms tested (6 research teams + 1 open-source)
- Infrastructure set up at MTG, Barcelona
- Data, annotations, scripts and individual results available
- `http://www.iua.upf.es/mtg/ismir2004/contest/tempoContest/`
- Gouyon et al. (2006)

OFAI

Data

- Preparatory data (no training data): 7 instances
- Test data: 3199 instances with tempo annotations ($24 < \text{BPM} < 242$)
- Linear PCM format, $>$ 12 hours
  - Loops: 2036 items, Electronic, Ambient, etc.
  - Ballroom: 698 items, Cha-Cha, Jive, etc.
  - Song excerpts: 465 items, Rock, Samba, Greek, etc.

Algorithms

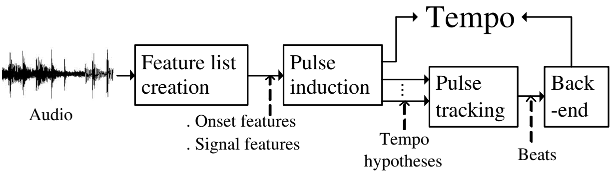


Figure: Tempo induction algorithms functional blocks

Algorithms

- Alonso et al. (2004): 2 algos
  - onsets
  - induction of 1 level by ACF or spectral product
  - tracking bypassed
- Dixon (2001a): 2 algos
  - onsets
  - IOI histogram
  - induction (+ tracking of 1 level + back-end)
- Dixon et al. (2003): 1 algo
  - energy in 8 freq. bands
  - induction of 2 levels by ACF
  - no tracking
- Klapuri et al. (2006): 1 algo
  - energy diff. in 36 freq. bands, combined into 4
  - comb filterbank
  - induction + tracking of 3 levels + back-end

Algorithms

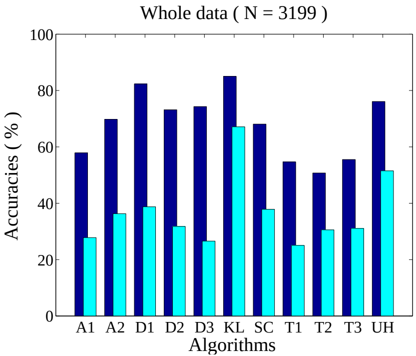
- Scheirer (1998): 1 algo `http://sound.media.mit.edu/~eds/beat/tapping.tar.gz`
  - energy diff. in 6 freq. bands
  - comb filterbank
  - induction + tracking of 1 level + back-end
- Tzanetakis and Cook (2002): 3 algos `http://www.sourceforge.net/projects/marsyas`
  - energy in 5 freq. bands
  - induction of 1 level by ACF
  - histogramming
- Uhle et al. (2004): 1 algo
  - energy diff. in freq. bands, combined in 1
  - induction of 3 level by ACF
  - histogramming

Evaluation Metrics

- Accuracy 1: Percentage of tempo estimates within 4% of ground-truth
- Accuracy 2: Percentage of tempo estimates within 4% of $1\times$, $\frac{1}{2}\times$, $\frac{1}{3}\times$, $2\times$ or $3\times$ ground-truth
- Width of precision window not crucial
- Test robustness against a set of distortions
- Statistical significance (i.e. McNemar test: errors on different instances $\Leftrightarrow$ significance)

Results

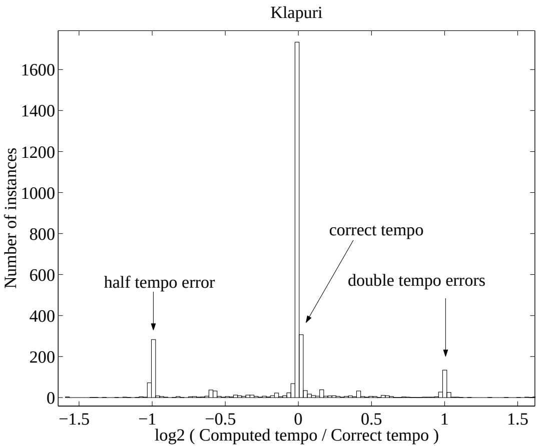
Figure: Accuracies 1 & 2



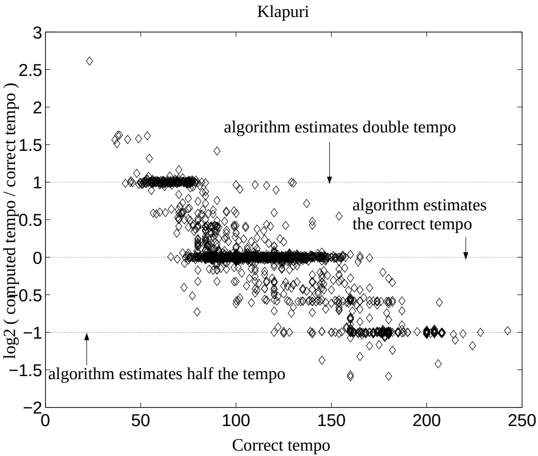
Results

- Klapuri et al. (2006) best on (almost) all data sets and metrics
- Accuracy 1: ˜63%
- Accuracy 2: ˜90%
- Clear tendency towards metrical level errors ($\Rightarrow$ Justification of Accuracy 2)
- Tempo induction feasible if we do not insist on a specific metrical level
- Worth of explicit moderate tempo tendency?
- Robust tempo induction $\Leftarrow$ frame features rather than onsets

Results

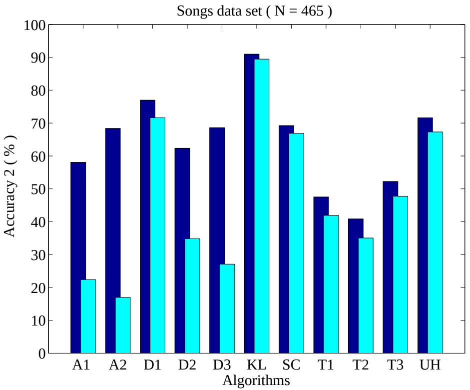

Klapuri
Number of instances
1600
1400
1200
1000
800
600
400
200
0
−1.5
−1
−0.5
0
0.5
1
1.5
log2 (Computed tempo / Correct tempo)
half tempo error
correct tempo
double tempo errors

Results


Klapuri
algorithm estimates double tempo
algorithm estimates
the correct tempo
algorithm estimates half the tempo
log2 (computed tempo / correct tempo)
Correct tempo
3
2.5
2
1.5
1
0.5
0
−0.5
−1
−1.5
−2
0
50
100
150
200
250

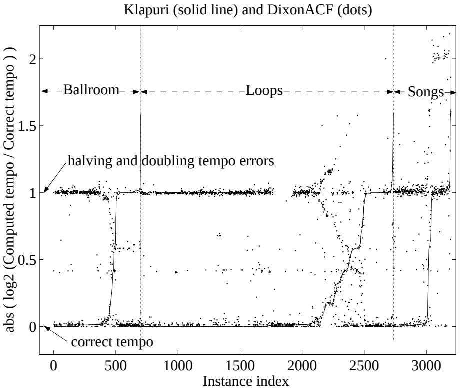
Results

Figure: Robustness test



Results

Figure: Errors on different items



Results

- Errors on different items
- Algorithms show unique performances on specific data
  - only 41 items correctly solved by *all algos*
  - 29 items correctly solved by a *single algo*
- Combinations better than single algorithms
  - median tempo does not work
  - voting mechanisms among “not too good” algorithms $\Rightarrow$ improvement
  - “Redundant approach”: multiple simple redundant mechanisms instead of a single complex algorithm (Bregman, 1998)
- Accuracy 2 requires knowledge of meter
- Ballroom data too “easy”
- Precision in annotations, more metadata

OFAI

Rhythm Classification - Outline

- Compare algorithms for automatic classification of 8 rhythm classes (Samba, Slow Waltz, Viennese Waltz, Tango, Cha Cha, Rumba, Jive, Quickstep) from audio data
- 1 algorithm (by Thomas Lidy et al.)
- Organisers did not enter the competition
- Data and annotations available
- `http://www.iua.upf.es/mtg/ismir2004/contest/rhythmContest/`

Data, Evaluations and Results

- 488 training instances
- 210 test instances
- Evaluation metrics: percentage of correctly classified instances
- Accuracy: 82%
- (see part on MIR applications)

Audio Tempo Extraction

- Proposed by Martin McKinney & Dirk Moelants at ISMIR 2005
- Task: “Perceptual tempo extraction”
- Tackling tempo ambiguity
  - different listeners may feel different metrical levels as the most salient
  - relatively ambiguous (61 or 122 BPM?)
    (courtesy of M. McKinney & D. Moelants)
  - relatively non-ambiguous (220 BPM)
    (courtesy of M. McKinney & D. Moelants)
  - assumption: this ambiguity depends on the signal
  - can we model this ambiguity?

Audio Tempo Extraction

- 13 algorithms tested (8 research teams)
- IMIRSEL infrastructure
- Evaluation scripts and training data available
- `http://www.music-ir.org/mirex2005/index.php/Audio_Tempo_Extraction`

Audio Tempo Extraction - Data

- Training data: 20 instances
- Beat annotated (1 level) by several listeners ($24 < N < 50$?) (Moelants and McKinney, 2004)
- Histogramming
- Derived metadata:
  - 2 most salient tempi
  - relative salience
  - phase first beat of each level
- Test data: 140 instances, same metadata

Audio Tempo Extraction - Algorithms

- Alonso et al. (2005): 1 algo
- Davies and Brossier (2005): 2 algos
- Eck (2005): 1 algo
- Gouyon and Dixon (2005a): 4 algos
- Peeters (2005): 1 algo
- Sethares (2005): 1 algo
- Tzanetakis (2005): 1 algo
- Uhle (2005): 2 algos

Audio Tempo Extraction - Evaluation Metrics

- Several tasks:
  - Task α: Identify most salient tempo (T1) within 8%
  - Task β: Identify 2nd most salient tempo (T2) within 8%
  - Task γ: Identify integer multiple/fraction of T1 within 8% (account for meter)
  - Task δ: Identify integer multiple/fraction of T2 within 8%
  - Task ϵ: Compute relative salience of T1
  - Task ζ: if α OK, identify T1 phase within 15%
  - Task η: if β OK, identify T2 phase within 15%
- $\forall$ tasks (apart ϵ) $\longleftarrow$ score 0 or 1
- $P = 0.25\alpha + 0.25\beta + 0.10\gamma + 0.10\delta$ $+0.20(1.0 - \frac{|\epsilon - \epsilon_{GT}|}{max(\epsilon, \epsilon_{GT})}) + 0.05\zeta + 0.05\eta$
- Statistical significance (McNemar)

Audio Tempo Extraction - Results

- `http://www.music-ir.org/evaluation/mirex-results/audio-tempo/index.html`
- Alonso et al. (2005) best P-score
- Some secondary metrics (on webpage, e.g. “At Least One Tempo Correct”, “Both Tempos Correct”)

Audio Tempo Extraction - Comments

- Very high standard deviations in performances
- Differences in performances not statistically significant
- Ranking from statistical test $\neq$ mean ranking
- Results on individual tasks not reported
  $\Rightarrow$ Individual results should be made public
- Task (modelling tempo ambiguity) is not representative of what competing algorithms really do (beat tracking or tempo induction at 1 level)
  $\Rightarrow$ Stimulate further research on tempo ambiguity
- Too many factors entering final performance
- "Tempo ambiguity modeling" contributes only 20% to final performance

Audio Tempo Extraction

- `http://www.music-ir.org/mirex2006/index.php/Audio_Tempo_Extraction`
- Simpler performance measure than MIREX 2005 (i.e. no phase consideration, no consideration of integer multiple/ratio of tempi)
- Thursday...

Audio Beat Tracking

- http://www.music-ir.org/mirex2006/index.php/Audio_Beat_Tracking
- Thursday...

More Benchmarks

- Rhythm patterns
- Meter
- Systematic deviations
- Quantisation
- etc.

Better Benchmarks

- Better data: more (and more accurate) annotations
- "Correct metrical level" problem
  - ISMIR04 data: too simple (no meter), MIREX05-06 data: too few (time-consuming annotations)
  - compromise: 1 single annotator per piece, annotations of two different levels, best match with algorithm output
  - assumption: two listeners would always agree on (at least) 1 level
- Richer metadata $\Rightarrow$ performance niches
  e.g. measuring "rhythmic difficulty" (Goto and Muraoka, 1997; Dixon, 2001b)
  - tempo changes
  - complexity of rhythmic patterns
  - timbral characteristics
  - syncopations

Better Benchmarks

- More modular evaluations
  - specific sub-measures (time precision, computational efficiency, etc.)
  - motivate submission of several variants of a system
- More open source algorithms
- Better robustness tests: e.g. increasing SNR, cropping
- Foster further analyses of published data ⇒ availability of:
  - data and annotations
  - evaluation scripts
  - individual results
- Statistical significance is a must (Flexer, 2006)
- Run systems several years (condition to entering contest?)

Part IV

Applications of Rhythm Description Systems

O F A I

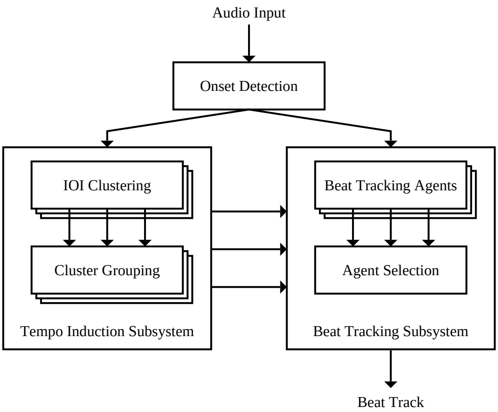
Outline

- MIR Applications
  - Interactive Beat Tracking
  - Audio Alignment
  - Classification with Rhythm Patterns
  - Query-by-Rhythm
- Rhythm Transformations
  - Tempo Transformations
  - Swing Transformations

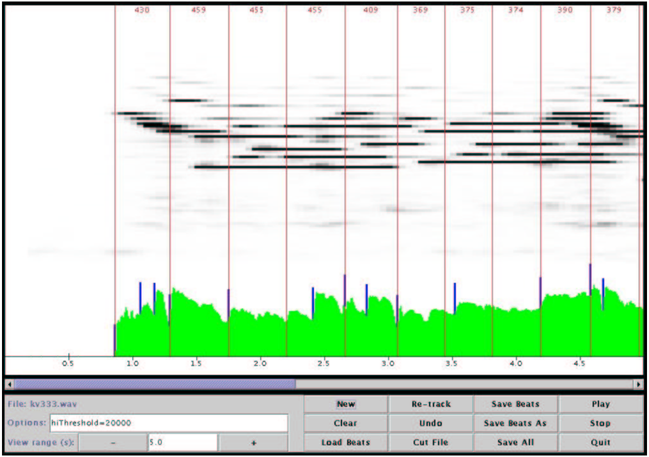
BeatRoot: Interactive Beat Tracking System

- Dixon (2001a,c)
- Annotation of audio data with beat times at various metrical levels
- Tempo and beat times are estimated automatically
- Interactive correction of errors with graphical interface
- New version available for download at: `http://www.ofai.at/~simon.dixon/beatroot`
  - improved onset detection (Dixon, 2006)
  - platform independent

BeatRoot Architecture



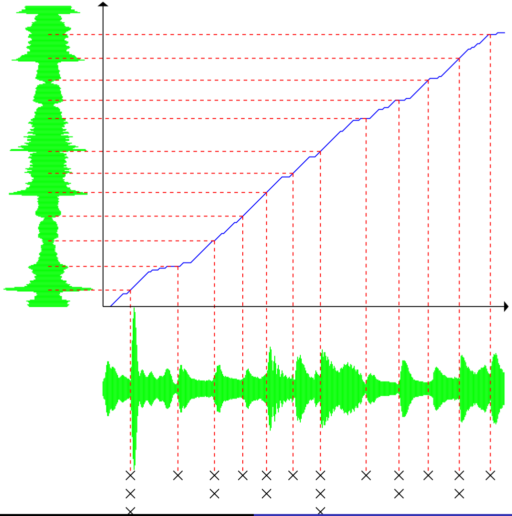
BeatRoot Demo



Audio Alignment

- Blind signal analysis is difficult
- Manual correction is tedious and error-prone
- In many situations, there is knowledge that is being ignored: e.g. the score, recordings of other performances, MIDI files
- Indirect annotation via audio alignment
  - Creates a mapping between the time axes of two performances
  - Content metadata from one performance can then be mapped to the other

Annotation via Audio Alignment

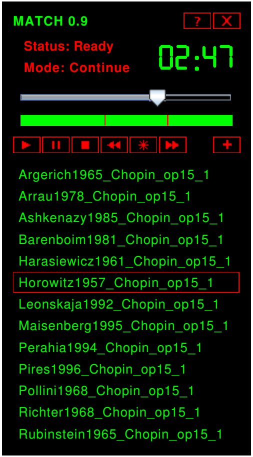


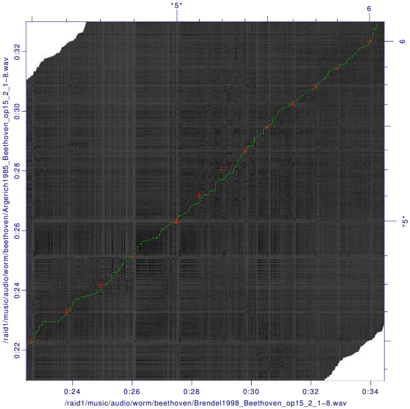
MATCH: Audio Alignment System

- Dixon (2005); Dixon and Widmer (2005)
- On-line time warping
  - linear time and space costs
  - robust real-time alignment
  - interactive interface
  - on-line visualisation of expression in musical performances
- How well does it work?
  - Off-line: average error 23ms on clean data
  - On-line: average error 59ms
  - Median error 20ms (1 frame)
- Available for download at:
  `http://www.ofai.at/~simon.dixon/match`

OFAI

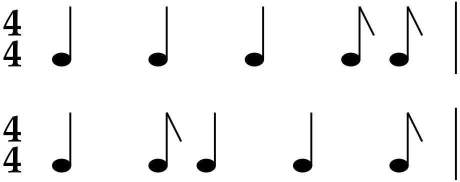
MATCH: Demo


MATCH 0.9
?
X
Status: Ready
Mode: Continue
02:47
Argerich1965_Chopin_op15_1
Arrau1978_Chopin_op15_1
Ashkenazy1985_Chopin_op15_1
Barenboim1981_Chopin_op15_1
Harasiewicz1961_Chopin_op15_1
Horowitz1957_Chopin_op15_1
Leonskaja1992_Chopin_op15_1
Maisenberg1995_Chopin_op15_1
Perahia1994_Chopin_op15_1
Pires1996_Chopin_op15_1
Pollini1968_Chopin_op15_1
Richter1968_Chopin_op15_1
Rubinstein1965_Chopin_op15_1


5
6
0:22
0:24
0:26
0:28
0:30
0:32
/raid1/music/audio/worm/beethoven/Argerich1985_Beethoven_op15_2_1-8.wav
0:24
0:26
0:28
0:30
0:32
0:34
/raid1/music/audio/worm/beethoven/Brendel1998_Beethoven_op15_2_1-8.wav

Classification with Rhythm Patterns

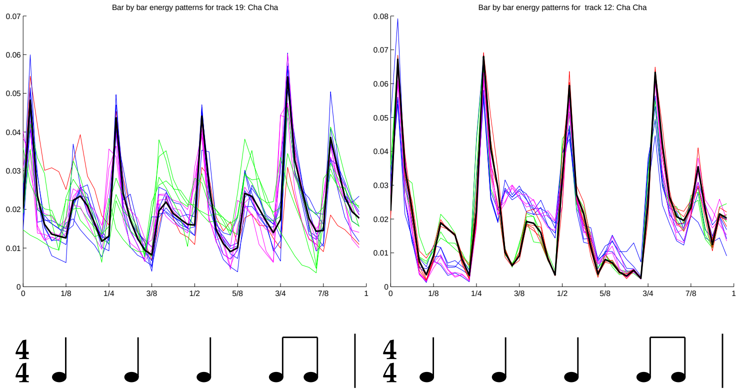
- Dixon et al. (2004)
- Classification of ballroom dance music by rhythm patterns
- Patterns: energy in bar-length segments
- One-dimensional vector
- Temporal order (within each bar) is preserved
- Musically meaningful interpretation of patterns (high level)



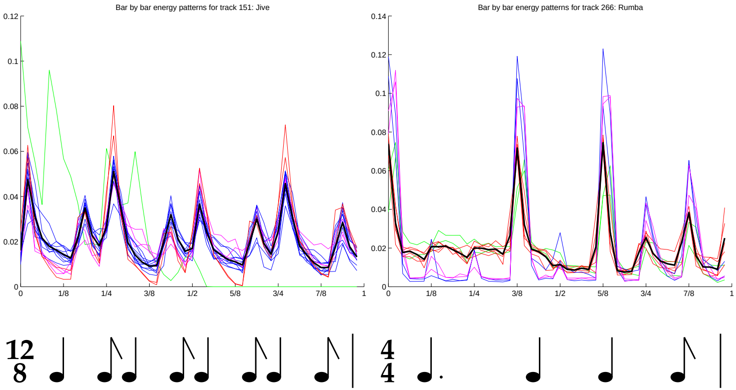
Pattern Extraction

- Tempo: BeatRoot and manual correction (first bar)
- Amplitude envelope: LPF & downsample
- Segmentation: correlation
- Clustering: k-means (k=4)
- Selection: largest cluster
- Comparison: Euclidean metric

Rhythm Pattern Examples: Cha Cha



More Rhythm Pattern Examples: Jive and Rumba



Classification

- Standard machine learning software: Weka
  - k-NN, J48, AdaBoost, Classification via Regression
- Feature vectors:
  - Rhythm pattern
  - Derived features
  - Periodicity histogram
  - IOI histogram / “MFCC”
  - Tempo

Classification Results

Feature sets	Without RP	With RP (72)
None (0)	15.9%	50.1%
Periodicity histograms (11)	59.9%	68.1%
IOI histograms (64)	80.8%	83.4%
Periodicity & IOI hist. (75)	82.2%	85.7%
Tempo attributes (3)	84.4%	87.1%
All (plus bar length) (79)	95.1%	96.0%

Discussion

- Only rhythm
  - No timbre (instrumentation), harmony, melody, lyrics
- One pattern
  - Sometimes trivial
- Short pieces (30 sec)
- Up to 96% classification

Query-by-Tapping

- Rhythm similarity computation between 2 symbolic sequences
  - Chen and Chen (1998); Peters et al. (2005)
  - `http://www.musipedia.org/query_by_tapping.0.html`
- Retrieving songs with same tempo as tapped query
  - Kapur et al. (2005)
  - `http://www.songtapper.com/`

Vocal queries (“Beat Boxing”)

- Kapur et al. (2004); Nakano et al. (2004); Gillet and Richard (2005a,b); Hazan (2005)

Query-by-Example

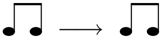
- Query = (computed) tempo
- Query = (computed) rhythm pattern (Chen and Chen, 1998; Kostek and Wojcik, 2005)
- Query = (computed) pattern + timbre data, e.g. drums (Paulus and Klapuri, 2002; Gillet and Richard, 2005b)

Synchronisation

- Applications to synchronisation of:
  - two audio streams
    - matching two streams in tempo and phase
    - done manually by DJ's
    - can be automated (Yamada et al., 1995; Cliff, 2000; Andersen, 2005) $\Rightarrow$ automatic sequencing in playlist generation
  - lights and music
    `http://staff.aist.go.jp/m.goto/PROJ/bts.html`

Tempo transformations

- Controlling tempo of audio signal (Bonada, 2000)



(courtesy of Jordi Bonada)

  - driven by gesture, conducting with infra-red baton, `www.hdm.at` (Borchers et al., 2002)
  - driven by tapping; secondary audio stream (Janer et al., 2006)

Swing transformations

- Delay of the 2nd, 4th, 6th & 8th eighth-note in a bar
- Example
  - eighth-notes
  - swinged eighth-notes
- Swing ratio
  - 2:1 ternary feel
  - depends on the tempo (Friberg and Sundström, 2002)

Acknowledgments: Lars Fabig & Jordi Bonada

OFAI

Swing transformation methods

- MIDI score matching
  - MIDI notes control the playback of mono samples
  - swing is added on MIDI
  - not suitable to polyphonic samples
  - sampler required
- Audio slicing (e.g. Recycle)
  - MIDI score controls playback of audio slices
  - same as above but samples are obtained from audio slices (can be polyphonic)
  - preprocessing:
    - slicing
    - mapping slices/MIDI notes
  - artificial tail synthesized on each slice $\Rightarrow$ sound quality $\downarrow$

Acknowledgments: Lars Fabig & Jordi Bonada

OFAI

"Swing transformer"

Gouyon et al. (2003)

- Similar to audio slicing but
  - no mapping necessary to MIDI
  - no artificial tail
  - use of time stretching algorithm
- Rhythmic analysis
  - onset detection
  - eighth-notes and quarter-notes period estimation
  - swing ratio estimation
  - eighth-notes and quarter-notes phase estimation
- Time stretching
  - odd eighth-notes are expanded
  - even eighth-notes are compressed

Acknowledgments: Lars Fabig & Jordi Bonada

“Swing transformer”

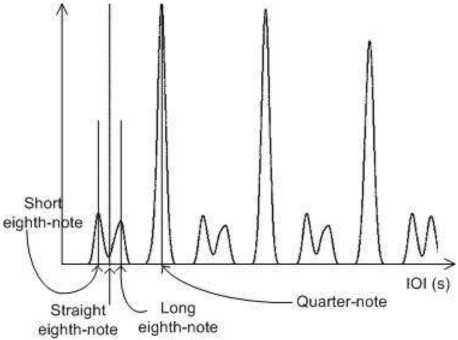


Figure: Swing ratio estimation

“Swing transformer”

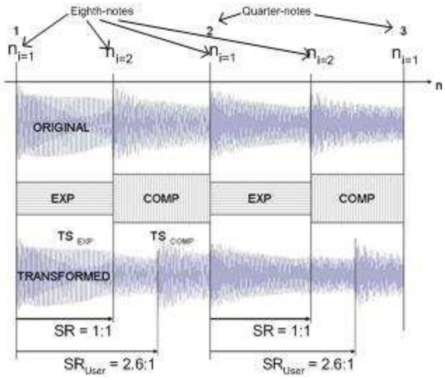
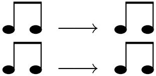


Figure: Expansion and compression of eighth-notes

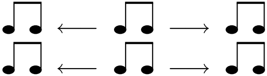
"Swing transformer"

Examples

- Add swing



- Add or remove swing



OFAI

Other rhythm transformations

- Automatic quantisation of audio
- Fit to a rhythm template
- Meter transformations: e.g. delete or repeat the last beat (Janer et al., 2006)
- Tempo- and beat-driven processing (Gouyon et al., 2002; Andersen, 2005)
- Concealment of transmission error in streamed audio by beat-based pattern matching (Wang, 2001)

Part V

Some ideas

Where Are We Going?

- Tempo induction, beat tracking, automatic transcription, genre recognition, melody extraction, etc.
  - all perform at $80 \pm 10\%$ accuracy
- The next step
  - Solve the next problem with 80% accuracy ??
  - Build better interfaces for interactive correction
  - Explore limitations of current approaches
- Limiting factors?
  - Computational power
  - Algorithms
  - Data
  - Knowledge: our models are too simple

Using Musical Knowledge

- What knowledge is available?
  - Data: score, recordings, MIDI files
  - Knowledge: music theory, performance "rules"
- What knowledge is relevant?
- Illustration: analysis of expressive timing in performance
  - Large project (since 1998)
  - Used beat tracking with manual correction to annotate recordings of famous pianists
  - Audio alignment: promises an order of magnitude decrease in work
  - What about high-level (musical) knowledge?

OFAI

Challenge: Encoding Musical Knowledge

- We don't know how to represent musical knowledge!
- Example 1: Machine learning of relationships between score and performance data
  - What are the relevant concepts? (Phrase structure, harmonic structure, etc)
  - How can these be computed?
- Example 2: Symbolic encoding of rhythm patterns for indexing and retrieval
  - One-dimensional energy patterns are limited
  - Multidimensional patterns would be better
  - e.g. frequency bands, instrumentation, drum sounds
  - Similarity metrics??

OFAI

Challenge: Modelling Low Levels of Perception

- Best low-level features for rhythm description?
- Different for different purposes (e.g. identifying beats, determining meter)?
- Different for different categories, music pieces?
- Consider more (and more high-level) features
  - auditory nerve cell firing models
  - pitch
  - chords
  - timbre
- Combine low-level features and onset features
- Deal with large numbers of features
- “Online” feature selection
- Perceptual validity of most efficient features

OFAI

Challenge: Modelling Low Levels of Perception

- "Redundant" approach (different simple low-level processes serve the same purpose)
  - which commonalities and differences (features, rhythm periodicity functions, etc.)?
  - how simple?
  - optimal voting scheme?
  - link with Ensemble Learning methods
- Tempo preference modelling
  - often implemented as scaling with curve centered around 120 BPM
  - evaluations showed artefacts for pieces with extreme tempo
  - modeling preference curve dependence on signal low-level attributes? which ones?
- Synchronisation in networks of simple rhythmic units, with acoustic inputs

Challenge: Observing Rhythm Perception

- Behavioral studies (Music Psychology, Neurophysiology of music)
- Different neural areas responsible for the perception of different rhythmic percepts? (Thaut, 2005)
  - high-level vs low-level processing
- Relations between imagined and perceived rhythm (Desain, 2004)
- Link between rhythm perception and rhythm production and motor control (Phillips-Silver and Trainor, 2005; Grahn and Brett, under review)

Filling In Gaps

- Methodological gap: Link observations and models
  - Ideally computational and behavioral methods should provide hypotheses and validation tools to each other
  - discrepancy in level of detail
- Semantic gap (lack of coincidence between algorithm representations and user interpretations): Which representations are meaningful?

Filling In Gaps

- Processing gap: Suitable processing architecture for combining top-down and bottom-up information flows?
- "Evolutional" gap: What is the purpose of our ability to perceive rhythm and what does perceiving rhythm share with e.g. cognition, speech, motor control?
  - sensory-motor theory of cognition
  - active rhythm perception. Explore link between rhythm perception and production by implementing rhythm perception modules on mobile robots immersed in musical environments (Brooks, 1991; Bryson, 1992)

OFAI

Part VI

Bibliography

Bibliography I

M. Alghoniemy and A. Tewfik. Rhythm and periodicity detection in polyphonic music. In *Proc. IEEE Workshop on Multimedia Signal Processing*, pages 185–190. IEEE Signal Processing Society, 1999.

P. Allen and R. Dannenberg. Tracking musical beats in real time. In *Proceedings of the International Computer Music Conference*, pages 140–143, San Francisco CA, 1990. International Computer Music Association.

Bibliography II

M. Alonso, B. David, and G. Richard. Tempo and beat estimation of musical signals. In *Proc. International Conference on Music Information Retrieval*, pages 158–163, Barcelona, 2004. Audiovisual Institute, Universitat Pompeu Fabra.

M. Alonso, B. David, and G. Richard. Tempo extraction for audio recordings. In *Proc. International Conference on Music Information Retrieval, MIREX posters*, 2005.

T. H. Andersen. *Interaction with Sound and Pre-recorded Music: Novel Interfaces and Use Patterns*. PhD Thesis, Department of Computer Science, University of Copenhagen, 2005.

Bibliography III

J. Bello. *Towards the Automated Analysis of Simple Polyphonic Music: A Knowledge-based Approach*. PhD Thesis, Dept of Electronic Engineering, Queen Mary University of London, London, 2003.

T. Blum, D. Keislar, A. Wheaton, and E. Wold. Method and article of manufacture for content-based analysis, storage, retrieval, and segmentation of audio information. USA Patent 5,918,223, 1999.

J. Bonada. Automatic technique in frequency domain for near-lossless time-scale modification of audio. In *Proc. of the International Computer Music Conference*, pages 396–399, 2000.

OFAI

Bibliography IV

J. O. Borchers, W. Samminger, and M. Mühlhäuser. Engineering a realistic real-time conducting system for the audio/video rendering of a real orchestra. In *IEEE MSE 2002 Fourth International Symposium on Multimedia Software Engineering*, 2002.

A. Bregman. Psychological data and computational ASA. In D. Rosenthal and H. Okuno, editors, *Computational Auditory Scene Analysis*. Lawrence Erlbaum Associates, New Jersey, 1998.

R. Brooks. New approaches to robotics. *Science*, 253: 1227–1232, 1991.

OFAI

Bibliography V

J. Brown. Determination of the meter of musical scores by autocorrelation. *Journal of the Acoustical Society of America*, 94(4):1953–1957, 1993.

J. Brown and M. Puckette. Calculation of a narrowed autocorrelation function. *Journal of the Acoustical Society of America*, 85(4):1595–1601, 1989.

J. Bryson. *The Subsumption Strategy Development of a Music Modelling System*. Master Thesis, University of Edinburgh, Faculty of Science, Department of Artificial Intelligence, Edinburgh, 1992.

OFAI

Bibliography VI

P. Cano, E. Gómez, F. Gouyon, P. Herrera, M. Koppenberger, B. Ong, X. Serra, S. Streich, and N. Wack. ISMIR 2004 audio description contest. MTG Technical Report: MTG-TR-2006-02, 2006.

A. Cemgil, P. Desain, and B. Kappen. Rhythm quantization for transcription. *Computer Music Journal*, 24(2):60–76, 2000a.

A. Cemgil and B. Kappen. Tempo tracking and rhythm quantisation by sequential monte carlo. In *Advances in Neural Information Processing Systems*, 2001.

A. Cemgil and B. Kappen. Monte carlo methods for tempo tracking and rhythm quantization. *Journal of Artificial Intelligence Research*, 18:45–81, 2003.

OFAI

Bibliography VII

A. Cemgil, B. Kappen, P. Desain, and H. Honing. On tempo tracking: Tempogram representation and Kalman filtering. In *Proceedings of the International Computer Music Conference*, pages 352–355, San Francisco CA, 2000b. International Computer Music Association.

A. Cemgil, B. Kappen, P. Desain, and H. Honing. On tempo tracking: Tempogram representation and Kalman filtering. *Journal of New Music Research*, 28(4):259–273, 2001.

J. Chen and A. Chen. Query by rhythm: An approach for song retrieval in music databases. In *Proceedings of the 8th IEEE International Workshop on Research Issues in Data Engineering*, pages 139–146, 1998.

Bibliography VIII

J. Chowning, L. Rush, B. Mont-Reynaud, C. Chafe, A. Schloss, and J. Smith. Intelligent system for the analysis of digitized acoustic signals. Report STAN-M-15, CCRMA, Stanford University, Palo Alto, 1984.

J. Chung. An agency for the perception of musical beats, or if i had a foot... Master's thesis, Massachusetts Institute of Technology, Cambridge, MA, 1989.

D. Cliff. Hang the DJ: Automatic sequencing and seamless mixing of dance-music tracks. Report HPL-2000-104 Hewlett Packard, 2000.

G. Cooper and L. Meyer. *The Rhythmic Structure of Music*. University of Chicago Press, Chicago, IL, 1960.

Bibliography IX

R. Dannenberg and B. Mont-Reynaud. Following an improvisation in real time. In *Proceedings of the International Computer Music Conference*, pages 241–258. International Computer Music Association, 1987.

M. Davies and P. Brossier. Fast implementations for perceptual tempo extraction. In *Proc. International Conference on Music Information Retrieval, MIREX posters*, 2005.

M. Davies and M. Plumbley. Beat tracking with a two state model. In *Proc. IEEE International Conference on Audio, Speech and Signal Processing*, 2005.

Bibliography X

P. Desain. What rhythm do I have in mind? detection of imagined temporal patterns from single trial ERP. In *Proc. 8th International Conference on Music Perception and Cognition*, page 209, 2004.

P. Desain and S. de Vos. Autocorrelation and the study of musical expression. In *Proc. International Computer Music Conference*, pages 357–360, 1990.

P. Desain and H. Honing. The quantization of musical time: A connectionist approach. *Computer Music Journal*, 13(3): 55–66, 1989.

Bibliography XI

P. Desain and H. Honing. Tempo curves considered harmful: A critical review of the representation of timing in computer music. In *Proceedings of the International Computer Music Conference*, 1991a.

P. Desain and H. Honing. Towards a calculus for expressive timing in music. *Computers in Music Research*, 3:43–120, 1991b.

P. Desain and H. Honing. Computational models of beat induction: The rule-based approach. *Journal of New Music Research*, 28(1):29–42, 1999.

S. Dixon. Automatic extraction of tempo and beat from expressive performances. *Journal of New Music Research*, 30(1):39–58, 2001a.

Bibliography XII

S. Dixon. An empirical comparison of tempo trackers. In *Proc. 8th Brazilian Symposium on Computer Music*, pages 832–840, 2001b.

S. Dixon. An interactive beat tracking and visualisation system. In *Proceedings of the International Computer Music Conference*, pages 215–218, 2001c.

S. Dixon. Live tracking of musical performances using on-line time warping. In *Proceedings of the 8th International Conference on Digital Audio Effects*, pages 92–97, 2005.

S. Dixon. Onset detection revisited. In *Proceedings of the 9th International Conference on Digital Audio Effects*, 2006.

OFAI

Bibliography XIII

S. Dixon and E. Cambouropoulos. Beat tracking with musical knowledge. In *Proc. European Conference on Artificial Intelligence*, pages 626–630. IOS Press, Amsterdam, 2000.

S. Dixon, W. Goebl, and E. Cambouropoulos. Perceptual smoothness of tempo in expressively performed music. *Music Perception*, 23(3):195–214, 2006.

S. Dixon, F. Gouyon, and G. Widmer. Towards characterisation of music via rhythmic patterns. In *5th International Conference on Music Information Retrieval*, pages 509–516, 2004.

S. Dixon, E. Pampalk, and G. Widmer. Classification of dance music by periodicity patterns. In *4th International Conference on Music Information Retrieval*, pages 159–165, 2003.

OFAI

Bibliography XIV

S. Dixon and G. Widmer. MATCH: A music alignment tool chest. In *6th International Conference on Music Information Retrieval*, pages 492–497, 2005.

J. Downie, editor. *The MIR/MDL evaluation project white paper collection*. Proceedings of ISMIR International Conference on Music Information Retrieval, 2002.

D. Eck. *Meter Through Synchrony: Processing Rhythmical Patterns with Relaxation Oscillators*. PhD thesis, Indiana University, Department of Computer Science, 2000.

D. Eck. A tempo-extraction algorithm using an autocorrelation phase matrix and shannon entropy. In *Proc. International Conference on Music Information Retrieval, MIREX posters*, 2005.

Bibliography XV

D. Eck. Finding long-timescale musical structure with an autocorrelation phase matrix. *Music Perception*, in press.

A. Flexer. Statistical evaluation of music information retrieval experiments. *Journal of New Music Research*, 35(2): 113–120, 2006.

J. Foote, M. Cooper, and U. Nam. Audio retrieval by rhythmic similarity. In *Proc. International Conference on Music Information Retrieval*, pages 265–266, 2002.

J. Foote and S. Uchihashi. The Beat Spectrum: A new approach to rhythm analysis. In *Proc. International Conference on Multimedia and Expo*, pages 881–884, 2001.

OFAI

Bibliography XVI

A. Friberg and J. Sundström. Swing ratios and ensemble timing in jazz performances: Evidence for a common rhythmic pattern. *Music Perception*, 19(3):333–349, 2002.

M. Gasser, D. Eck, and R. Port. Meter as mechanism: A neural network that learns metrical patterns. *Connection Science*, 11(2):187–216, 1999.

O. Gillet and G. Richard. Drum loops retrieval from spoken queries. *Journal of Intelligent Information Systems*, 24(2/3): 159–177, 2005a.

O. Gillet and G. Richard. Indexing and querying drum loops databases. In *Proceedings of the Fourth International Workshop on Content-Based Multimedia Indexing*, 2005b.

OFAI

Bibliography XVII

E. Gómez. *Tonal Description of Music Audio Signals*. PhD Thesis, Universitat Pompeu Fabra, Barcelona, 2006.

M. Goto and Y. Muraoka. A real-time beat tracking system for audio signals. In *Proceedings of the International Computer Music Conference*, pages 171–174, San Francisco CA, 1995. International Computer Music Association.

M. Goto and Y. Muraoka. Issues in evaluating beat-tracking systems. In *Proc. International Joint Conferences on Artificial Intelligence, Workshop on Computational Auditory Scene Analysis*, pages 9–16, 1997.

M. Goto and Y. Muraoka. Real-time beat tracking for drumless audio signals. *Speech Communication*, 27(3–4):311–335, 1999.

O F A I

Bibliography XVIII

F. Gouyon. *Extraction automatique de descripteurs rythmiques dans des extraits de musiques populaires polyphoniques*. Master's Thesis & Internal report, IRCAM, Centre Georges Pompidou, Paris & Sony CSL, Paris, 2000.

F. Gouyon and S. Dixon. Influence of input features in perceptual tempo induction. In *Proc. International Conference on Music Information Retrieval, MIREX posters*, 2005a.

F. Gouyon and S. Dixon. A review of automatic rhythm description systems. *Computer Music Journal*, 29(1):34–54, 2005b.

Bibliography XIX

F. Gouyon, S. Dixon, E. Pampalk, and G. Widmer. Evaluating rhythmic descriptors for musical genre classification. In *Proc. 25th International AES Conference*, pages 196–204, London, 2004. Audio Engineering Society.

F. Gouyon, L. Fabig, and J. Bonada. Rhythmic expressiveness transformations of audio recordings: Swing modifications. In *Proc. International Conference on Digital Audio Effects*, pages 94–99, London, 2003.

F. Gouyon and P. Herrera. A beat induction method for musical audio signals. In *Proc. 4th European Workshop on Image Analysis for Multimedia Interactive Services*, pages 281–287, Singapore, 2003a.

O F A I

Bibliography XX

F. Gouyon and P. Herrera. Determination of the meter of musical audio signals: Seeking recurrences in beat segment descriptors. In *Proc. 114th AES Convention*. Audio Engineering Society, 2003b.

F. Gouyon, P. Herrera, and P. Cano. Pulse-dependent analyses of percussive music. In *Proc. 22nd International AES Conference*, pages 396–401. Audio Engineering Society, 2002.

F. Gouyon, A. Klapuri, S. Dixon, M. Alonso, G. Tzanetakis, C. Uhle, and P. Cano. An experimental comparison of audio tempo induction algorithms. *IEEE Trans. Speech and Audio Processing*, 14(5):1832–1844, 2006.

Bibliography XXI

F. Gouyon, G. Widmer, X. Serra, and A. Flexer. Acoustic cues to beat induction: A machine learning perspective. *Music Perception*, in press.

J. Grahn and M. Brett. Rhythm perception in motor areas of the brain. under review.

A. Hazan. Towards automatic transcription of expressive oral percussive performances. In *Proceedings of International Conference on Intelligent User Interfaces*, San Diego, CA, USA, 2005.

H. Honing. Issues in the representation of time and structure in music. *Contemporary Music Review*, 9:221–239, 1993.

OFAI

Bibliography XXII

H. Honing. From time to time: The representation of timing and tempo. *Computer Music Journal*, 25(3):50–61, 2001.

J. Janer, J. Bonada, and S. Jordà. Groovator – An implementation of real-time rhythm transformations. In *Proceedings of 121st Convention of the Audio Engineering Society*, San Francisco, CA, USA, 2006.

A. Kapur, M. Benning, and G. Tzanetakis. Query-by-beat-boxing: Music retrieval for the DJ. In *Proceedings of the Fifth International Conference on Music Information Retrieval*, 2004.

A. Kapur, R. McWalter, and G. Tzanetakis. New interfaces for rhythm-based retrieval. In *Proceedings of the International Conference on Music Information Retrieval*, 2005.

OFAI

Bibliography XXIII

A. Klapuri. Sound onset detection by applying psychoacoustic knowledge. In *Proc. IEEE International Conference on Acoustics, Speech and Signal Processing*, 1999.

A. Klapuri. Automatic music transcription as we know it today. *Journal of New Music Research*, 33(3):269–282, 2004.

A. Klapuri, A. Eronen, and J. Astola. Analysis of the meter of acoustic musical signals. *IEEE Transactions on Speech and Audio Processing*, 14(1), 2006.

B. Kostek and J. Wojcik. Automatic retrieval of musical rhythmic patterns. In *Proc. 119th AES Convention*, 2005.

E. Large and J. Kolen. Resonance and the perception of musical meter. *Connection Science*, 6:177–208, 1994.

Bibliography XXIV

J. Laroche. Estimating tempo, swing and beat locations in audio recordings. In *Proc. IEEE Workshop on Applications of Signal ‘ Processing to Audio and Acoustics*, pages 135–138, 2001.

J. Laroche. Efficient tempo and beat tracking in audio recordings. *Journal of the Acoustical Society of America*, 51 (4):226–233, 2003.

F. Lerdahl and R. Jackendoff. *A Generative Theory of Tonal Music*. MIT Press, Cambridge MA, 1983.

Bibliography XXV

T. Lidy and A. Rauber. Evaluation of feature extractors and psycho-acoustic transformations for music genre classification. In *Proceedings of ISMIR International Conference on Music Information Retrieval*, pages 34–41, London, 2005.

H. Longuet-Higgins. *Mental processes*. MIT Press, Cambridge, 1987.

H. Longuet-Higgins and C. Lee. The perception of musical rhythms. *Perception*, 11:115–128, 1982.

H. Longuet-Higgins and C. Lee. The rhythmic interpretation of monophonic music. *Music Perception*, 1(4):424–441, 1984.

O F A I

Bibliography XXVI

J. Maher and J. Beauchamp. Fundamental frequency estimation of musical signals using a two-way mismatch procedure. *Journal of the Acoust. Soc. of America*, 95(4): 2254–2263, 1993.

J. McAuley. *Perception of Time as Phase: Towards an Adaptive-Oscillator Model of Rhythmic Pattern Processing*. PhD thesis, Indiana University, Bloomington, 1995.

M. McKinney and J. Breebaart. Features for audio and music classification. In *4th International Conference on Music Information Retrieval*, pages 151–158, 2003.

M. McKinney and D. Moelants. Extracting the perceptual tempo from music. In *Proc. International Conference on Music Information Retrieval*, 2004.

OFAI

Bibliography XXVII

B. Meudic. Automatic meter extraction from MIDI files. In *Proc. Journées d'informatique musicale*, 2002.

D. Moelants. Preferred tempo reconsidered. In *Proc. of the International Conference on Music Perception and Cognition*, pages 580–583, Sydney, 2002.

D. Moelants and M. McKinney. Tempo perception and musical content: What makes a piece fast, slow or temporally ambiguous. In *Proc. International Conference on Music Perception and Cognition*, 2004.

T. Nakano, J. Ogata, M. Goto, and Y. Hiraga. A drum pattern retrieval method by voice percussion. In *Proc. International Conference on Music Information Retrieval*, 2004.

OFAI

Bibliography XXVIII

E. Pampalk. *Computational Models of Music Similarity and their Application to Music Information Retrieval*. Phd Thesis, Vienna University of Technology, Vienna, 2006.

E. Pampalk, S. Dixon, and W. G. Exploring music collections by browsing different views. In *Proc. International Conference on Music Information Retrieval*, pages 201–208, 2003.

E. Pampalk, A. Rauber, and D. Merkl. Content-based organization and visualization of music archives. In *Proc. ACM International Conference on Multimedia*, pages 570–579, 2002.

R. Parncutt. A perceptual model of pulse salience and metrical accent in musical rhythms. *Music Perception*, 11(4):409–464, 1994.

Bibliography XXIX

J. Paulus and A. Klapuri. Measuring the similarity of rhythmic patterns. In *Proceedings of the 3rd International Conference on Musical Information Retrieval*, pages 150–156. IRCAM Centre Pompidou, 2002.

G. Peeters. Tempo detection and beat marking for perceptual tempo induction. In *Proc. International Conference on Music Information Retrieval, MIREX posters*, 2005.

G. Peeters. Template-based estimation of time-varying tempo. *EURASIP Journal on Applied Signal Processing*, in press.

G. Peters, C. Anthony, and M. Schwartz. Song search and retrieval by tapping. In *Proc. Twentieth National Conference on Artificial Intelligence*, 2005.

O F A I

Bibliography XXX

J. Phillips-Silver and L. Trainor. Feeling the beat: Movement influences infant rhythm perception. *Science*, 308(5727): 1430, 2005.

D. Povel and P. Essens. Perception of temporal patterns. *Music Perception*, 2(4):411–440, 1985.

C. Raphael. A hybrid graphical model for rhythmic parsing. *Artificial Intelligence*, 137(1), 2002.

B. Repp. On determining the basic tempo of an expressive music performance. *Psychology of Music*, 22:157–167, 1994.

D. Rosenthal. Emulation of human rhythm perception. *Computer Music Journal*, 16(1):64–76, 1992a.

Bibliography XXXI

D. Rosenthal. *Machine rhythm: Computer emulation of human rhythm perception*. PhD Thesis, MIT, Cambridge, 1992b.

R. Rowe. Machine listening and composing with Cypher. *Computer Music Journal*, 16(1):43–63, 1992.

E. Scheirer. Pulse tracking with a pitch tracker. In *Proc. IEEE Worshop on Applications of Signal Processing to Audio and Acoustics*, 1997.

E. Scheirer. Tempo and beat analysis of acoustic musical signals. *Journal of the Acoustical Society of America*, 103(1): 588–601, 1998.

E. Scheirer. *Music-listening systems*. PhD Thesis, MIT, Cambridge, 2000.

Bibliography XXXII

E. Scheirer and M. Slaney. Construction and evaluation of a robust multifeature speech/music discriminator. In *Proc. IEEE-ICASSP*, pages 1331–1334, 1997.

A. Schloss. *On the automatic transcription of percussive music - From acoustic signal to high-level analysis*. PhD Thesis, CCRMA, Stanford University, Palo Alto, 1985.

J. Seppänen. *Computational models of musical meter recognition*. Master's Thesis, Tampere University of Technology, Tampere, 2001a.

J. Seppänen. Tatum grid analysis of musical signals. In *Proc. IEEE Workshop on Applications of Signal Processing to Audio and Acoustics*, 2001b.

OFAI

Bibliography XXXIII

W. Sethares. Tempo extraction via the periodicity transform. In *Proc. International Conference on Music Information Retrieval, MIREX posters*, 2005.

W. Sethares, R. Morris, and J. Sethares. Beat tracking of musical performances using low-level audio features. *IEEE Trans. Speech and Audio Processing*, 13(2):275–285, 2005.

W. Sethares and T. Staley. Meter and periodicity in musical performance. *Journal of New Music Research*, 30(2): 149–158, 2001.

L. Smith. Modelling rhythm perception by continuous time-frequency analysis. In *Proc. International Computer Music Conference*, 1996.

Bibliography XXXIV

D. Temperley. An evaluation system for metrical models. *Computer Music Journal*, 28(3):28–44, 2004.

D. Temperley and D. Sleator. Modeling meter and harmony: A preference rule approach. *Computer Music Journal*, 23(1): 10–27, 1999.

M. Thaut. *Rhythm, Music, and the Brain; Scientific Foundations and Clinical Applications*. Routledge, NY, 2005.

P. Todd, C. Lee, and D. Boyle. A sensory-motor theory of beat induction. *Psychological Research*, 66(1):26–39, 2002.

G. Tzanetakis. Tempo extraction using beat histograms. In *Proc. International Conference on Music Information Retrieval, MIREX posters*, 2005.

OFAI

Bibliography XXXV

G. Tzanetakis and P. Cook. Musical genre classification of audio signals. *IEEE Transactions on Speech and Audio Processing*, 10(5):293–302, 2002.

C. Uhle. Tempo induction by investigating the metrical structure of music using a periodicity signal that relates to the tatum period. In *Proc. International Conference on Music Information Retrieval, MIREX posters*, 2005.

C. Uhle, J. Rohden, M. Cremer, and J. Herre. Low complexity musical meter estimation from polyphonic music. In *Proc. AES 25th International Conference*, pages 63–68. Audio Engineering Society, 2004.

Bibliography XXXVI

L. van Noorden and D. Moelants. Resonance in the perception of musical pulse. *Journal of New Music Research*, 28(1): 43–66, 1999.

B. Vercoe. Computational auditory pathways to music understanding. In I. Deliège and J. Sloboda, editors, *Perception and Cognition of Music*, pages 307–326. Psychology Press, 1997.

Y. Wang. A beat-pattern based error concealment scheme for music delivery with burst packet loss. In *Proc. International Conference on Multimedia and Expo*, 2001.

E. Wold, T. Blum, D. Keislar, and J. Wheaton. Classification, search, and retrieval of audio. In *CRC Handbook of Multimedia Computing*. 1999.

OFAI

Bibliography XXXVII

Y. Yamada, T. Kimura, T. Funada, and G. Inoshita. An apparatus for detecting the number of beats. USA Patent 5,614,687, 1995.

O F A I