Beat Induction and Rhythm Analysis for Live Audio Processing: 1st Year PhD Report

Nick Collins[1]
Centre for Music and Science
Music Faculty
University of Cambridge
11 West Road
Cambridge
CB3 9DP
UK
http://www.cus.cam.ac.uk/ nc272/

June 18, 2004

[1]nc272@cam.ac.uk

Contents

Chapter 1

Introduction

Advanced research in computer music is driven by a desire to understand and match human abilities in the perception and production of audio events. Such an effort is essential to building automated transcription systems, sensitive computer accompanists and electronic improvisation partners. Cognitive science provides a vast literature of great relevance to these tasks, and much good work already exists across the fields of engineering, artificial intelligence, music theory, experimental psychology and neuroscience. Solutions for real-time work remain elusive, however, and an open topic of investigation; this thesis is a practical attempt to build a working system for concert use. Whilst cognitive science will provide an inspiration and grounding for what follows, an engineering solution may require a pragmatic approach to the implementation, and the solutions obtained herein may be less than general, indeed, slanted to the compositional needs of the interactive system under development. That particular target is an existing technology for live audio cutting detailed later in this report.

This thesis will describe attempts to build a real-time interactive system for computer music which allows a human instrumentalist to establish the rhythmic frame of performance. The computer must search for evidence of metrical hierarchy or timeline from the audio signal, tracking tactus (beat induction) whilst allowing for expressive timing and motor noise, and all this without excessive zeal but within some reasonably smooth trajectory.

For this research report, the review sections below cover interaction (1.1), the psychology of music as it pertains to rhythm (2.1) with a further selection of results immediately pertinent to beat induction (2.2). The third chapter is enclosed as additional evidence of review work, covering beat induction models from various fields (3.1) and a critique of beat induction itself (3.2). Note that for the shared bibliography it appears as part of this report, yet in word count should be taken as a separate example chapter, and it would be further expanded for a final thesis. A final chapter discusses the proposed model (4.1); for this public version, the research timetable has been removed to preserve my dignity as the inevitable twists and turns of PhD research take place!

1.1 Interaction Research in Computer Music

The computer provides many possibilities as an improvising partner. Histories and reviews are provided in [Roa96, chapters 14 and 15] [Dea03, Imp01, Jor02, Row93, Row01]. Yet too many electronic music pieces involve no knowledge in the computer accompaniment of the expressive timing of the human performer. This is no doubt because of the extremely non-trivial task of machine listening.

1.1.1 Accompaniment

For fixed score pieces, the naive historical approach is simple playback of a tape with the onus on the human performer to sync up, a case still seen at most current electroacoustic music festivals. A slight relaxation is a cue system where an operator will trigger sound events manually to match critical points in the score [MSWW01]. The automation of accompaniment is a natural yet tricky step. For monophonic instruments (especially the flute), this has been achieved by pitch tracking algorithms and pattern matching against a score. Barry Vercoe's *Synthetic Performer* was an early system (1984) developed at IRCAM, used for tracking flute and violin; it took advantage of fingering information from sensors as well as the audio stream. Dannenberg reviews early attempts at accompaniment [Dan89]. For robustness in the face of expressive timing and performer error, Chris Raphael's Bayesian system Music

Plus One [Rap01b, Rap04] is impressive.[1] Polyphonic tracking is extremely difficult, but some success is possible simply by using features of the audio stream (like spectral density), rather than attempting a complete ongoing transcription solution [JMF02]. The state of the art systems use statistical pattern matching to synchronise accompaniment with the most likely score location [OD01, PB02, Rap04].

1.1.2 Interaction with Improvisation

The degree of improvisation in performance is actually a continuum from predetermined structure[2] to full freedom of parameters. An intermediate case, where the known structure of a jazz chord sheet supports the improvisation of a solo, is tackled by Toiviainen's jazz accompanist [Toi98], the twist being that the accompanist has a repertoire of standards and will try to match the standard it must accompany during a lead-in by the soloist. Toiviainen makes use of oscillator based beat induction and Bayesian inference, and his research is further discussed as a technical inspiration below.

Cognitive aspects of improvisation were studied by Johnson-Laird amongst others [JL91]. He notes that

> The conjecture that modern jazz rhythms are generated by processes that place a minimal load on working memory proves to be borne out, and it should be possible to characterise the complete set of such phrases using a regular grammar. [JL91, page 305]

He was also the first to consider the problem of over or under generation of material from a database. Jazz has been a popular target for (sometimes) interactive algorithmic composition packages. An AI implementation of a jazz bass player which can selectively reuse phrase segments is presented in [RRG99]. From a HCI perspective, Walker devises a computer improvisor based on principles of conversation analysis [Wal97]. Thom trains her unsupervised learning system *Band-OUT-of-a-Box* on Charlie Parker solos [Tho00]. Bile's *GENJAM* [Bil02] is a genetic algorithm based live performance system which has been extensively tested in real concerts, as a practising musician in his Virtual Quintet (http://www.it.rit.edu/ jab/GenJam.html). The archetypical algorithmic composition of solos in jazz occurs in the commercial product Band-in-a Box (http://www.pgmusic.com/), which provides canned generative accompaniment but does not analyse the input from a human partner for expressive tempo variation or material.

More abstract interaction systems investigate the potential for new music in computerised systems. Robert Rowe's *Cypher* is a relatively general set of machine listening functions and is discussed in detail in [Row93, Row01]; Rowe also provides some useful history and context. Jonathan Impett's system [Imp01] is based on emergence, the appearance of significant global behaviour from the interaction of a multitude of more primitive agents. Impett plays an adapted trumpet, with a pitch to MIDI convertor. Another pitch to MIDI enabled musician is George Lewis, who recently received the MacArthur genius award, and whilst (as a predominantly practising musician) he has not published, has been interviewed and discussed by a number of authors including Dean, Roads and Casserley [Cas97]. His personal and unfortunately unexaminable *Voyager* software tracks his trombone playing and generates responses.

Much exploitation of computers in live performance makes use of the computer as a powerful effects unit, under the guidance of a human pilot rather than with any artificial intelligence techniques, capturing audio from the performer and processing it on the fly without necessarily analysing via computer the musical character of that audio. Softwares like STEIM's LiSa (Live Sampling) embody this, or Joel Ryan's work with Evan Parker combining the Eventide Harmonizer and SuperCollider 2.

The powerful and easily portable laptop makes live computer music an increasingly common experience in bars and clubs, though the great majority is not yet a profoundly interactive experience in the traditional sense. The mainstay are fixed commercial software packages for DJ mixing or sequence playback like Ableton Live, Traktor or Reason. More customisable softwares do exist (Reaktor, Max/MSP, PD), the most powerful being full programming languages specialised to audio (real-time Csound, SuperCollider, ChucK). The use of such generative and interactive software in laptop performance is discussed in [Col03].

1.1.3 New Musical Instruments

To many artists, electronic music provides fascinating opportunities to build novel controllers and new musical instruments, or expand the capabilities of traditional ones. Research institutes like STEIM are

[1] Raphael plays oboe: my Concerto for Accompaniment was written for him with his real-time computer accompanist on three pianos.

[2] The classical score is not the most rigid case, since it still allows expressive license.

dedicated to the investigation of new gestural and analytic interfaces [Rya91]. A new conference series NIME (New Interfaces for Musical Expression) is covering this exploration [Coo01].

As one examplar of this very practical research, Tod Machover's Brain Opera project at MIT is certainly noteworthy for the variety of novel interfaces explored [Par99]. A particularly interesting interface by this group (because of the club association) are the sensor laden dancing shoes [PHH99] which send 16 control streams covering such elements as elevation, acceleration, orientation and pressure.

The most ubiquitous controller is still the computer and its standard interface devices of keyboard and mouse. Networked music provides some new paradigms for interaction, with many online music systems and collaborative systems for musicians. Sergi Jorda's FMOL provides an outstanding case, and the system is also used in concert as a performance instrument [Jor02].

A recent reaction against the gestural heritage of musicianship as perhaps unnecessary for live computer music places the computer programming environment at the heart of performance. This is the domain of live coding or on-the-fly programming as an artistic activity [CMRW03, WC04]. An intriguing prototype, also much quoted as the first network band, is the Hub, a collective of musician programmers.

Chapter 2

Cognition of Rhythm

2.1 Rhythm, Metre and Pulse

Richard Parncutt reviews definitions of rhythm and settles upon one founded in his model of pulse salience:

> 'A musical rhythm is an acoustic sequence evoking a sensation of pulse'[1] [Par94, page 453]

This definition or something similar is accepted by most beat induction models which aim to find the intuitively natural 'foot-tapping' or 'hand-clapping' *tactus*, *referent time level* or *beat*, and the practical description adopted herein. Note that Parncutt's pulse salience refers to '*all* rhythmic levels spontaneously evoked in the mind of a listener' and that the beat is the comfortable middle ground of a metrical hierarchy.

The human experience of rhythm is not an exclusively Western phenomena, yet Western musical tradition places many weighted terms in the path of the analyst. Clayton posits

> 'Metre as commonly understood in the West is clearly not a universal concept, nor is it a phenomenon observable in all world musics' [Cla00, page 41]

He notes that the well-formedness rules for metre of Lerdahl and Jackendoff's theory [LJ83] cannot accommodate North Indian tāl patterns. The inadequacy of some of GTTM's rules as cultural universals is independently raised with respect to the music of the Bolivian campesinos by Cross [Cro99]. Temperley [Tem01], in his computational implementation of GTTM, revises some of the rules in a treatment of African rhythm, showing that the basic idea of well-formedness and preference rules is fruitful.

Yet there are at least three theories of metre for African rhythm. Arom [Aro89] finds an isoperiodic pulse level and subdividing operational value at the heart of Central African polyrhythm, rejecting though any sense of strong and weak accentuation within a cycle as arising from hierarchical metre. Agawu [Aga95] meanwhile argues for a conventional metrical backdrop to the Northern Ewe music of Ghana. Magill and Pressing [MP97] describe the nonisochronous *timeline* as the the best fit for a West African drummer's mental model of polyrhythmic production.[2]

Accepting for now that certain definitions of metre are not necessarily valid cross-culturally, common practise Western music remains the focus of most theory. Grouping (rhythmic phrasing) is separated from metre (the pulse hierarchy) in modern treatments[3]. Bruno Repp defines the rhythmicality of music as the degree to which it lends itself to division into perceptual groups and metricality as its degree of perceived temporal regularity [Rep00a, page 235]. In the Generative Theory of Tonal Music view of metre [LJ83], the hierarchy gives rise to strengths of metrical accentuation (the strong and weak beats of popular parlance) for measure locations based on how many metrical levels coincide at that point.

Resolving rhythm and metre can seem a chicken and egg problem: rhythm is understood from a metrical context but metre is inferred from the presented rhythm! This impasse may be overcome if a sense of metre is initialised from the first presented information, then itself sets up expectancies, with

[1] Thereby rendering certain contemporary composer's rhythms amusical, or at least redefining the perceptible musical effect of them. The definition may also make certain rhythms musical for those who are trained to hear!

[2] Metre through the imposition of time signature is not a necessity for all Western music; explicit barlines were introduced in the seventeenth century and disappear again in Faure and Satie piano pieces. Composers after the romantic period are influenced by wider conceptions of rhythm found in world musics, from Stravinsky and Bartok's use of additive metres through Messiaen's treatise and teaching. Ligeti's piano etude *Galamb Borong* is prefaced by the instruction 'the piece has no proper metre and the bar lines do not indicate any structure' [Lig98, page 4].

[3] Hasty gives an unclearly argued exception [Has97].

respect to which rhythms are interpreted. Desain and Honing [DH99] talk of a 'bottom-up' process establishing a sense of beat over 5-10 events, then a 'top-down' process operating to resolve rhythms. This is not the final say because 'when in a change of meter the evidence for the old percept becomes too meagre, a new beat interpretation is induced' [DH99, page 29]. From a Bayesian perspective there is a dual potential for the inference of one given any evidence of the other. Todd has written of the complementation of the two as being akin to a frequency to time domain transform [Cla99, page 478]; his rhythmogram produces images of rhythmic grouping structure from 'mexican hat' filtering of nerve firing signals [TB96], bringing together low-level and high-level representations.

Because of the characteristics of marking events and attentional mechanisms, rhythm is not purely a product of time location, and has been laid out as a multi-dimensional attribute by some authors. Though some studies based on inter onset intervals (IOIs) factor them out, timbre, envelope of attack and perceptual centre, fundamental frequency and amplitude all have a part to play in the tracking and segmenting of events. Many authors make this point (usually as a proviso to their IOI based studies) but the assistance of pitch information in resolving metrical levels is shown in [TS00].

Rhythms in human performance are not clinical and metronomic. Aside from noise[4], they show structure specific timing deviations which are a basic part of musical expression. These are absolutely locked to the perception of deep structure [Gab99]. An example is the agogic accent, the lengthening of the duration of an important note. Researchers at the Music Acoustics Group at KTH, Stockholm (http://www.speech.kth.se/music/performance/) have been compiling a set of musical performance rules (currently around 30 distinct factors have been proposed) which are available in the Director Musices software and have recently been parametrised by emotional states [SFF91, CRZF03].

The extent of expressive timing is such that a notated quarter note in one bar can be more than twice the duration of a minim in another [DH92]. This makes the quantisation processes in automated score transcription require musical knowledge of local context. Desain and Honing also sound a note of caution for those who might apply tempo curves with timing deviations locked in proportionally: it is wrong to assume that such perturbations scale exactly to different tempi, and the function for timing deviation must arise from the interaction of structure with motor constraints and pulse preferences [DH94].

The 'close relationship between music and human movement' [Cla99, page 494] is corroborated by recent neural imaging studies[JG03, LM03, SFvC00]. Fraisse asserts 'all of the rhythms that we perceive are rhythms which originally resulted from human activity' [Fra82, page 150]. Neil Todd's model of musical expression in dynamics [Tod92] and in earlier papers for tempo, explicitly utilises a dynamical system based in kinematics. Todd finds that 'the sensitivity of the auditory system to tempo is coincident with the natural frequency of the motor system- perhaps a product of the coevolution of the auditory and motor systems' [TB96, page 269], a point that Fraisse's review also makes from a wealth of supporting evidence; haptic motion is at rhythmic rates, and perception of beat prefers a foot-tapping tempo.

The physiology and neurology of the human auditory system pertaining to perception of temporal events is covered for instance in [EG97, Moo97b, VP93]. Computational auditory models use a direct embodiment of such research (for example [TLO02]), and most engineering attempts at beat induction have psychoacoustical foundations [Sch98]. The non-linearity of the human ear is worth remark: whilst linear time-frequency transforms are constrained by the Heisenberg Uncertainty Principle, the ear can exceed a linear system's resolution by a factor of five [Har97, page 28]. Also relevant is the cognition of events in complex auditory mixtures as occur in real environments, now commonly known as auditory scene analysis [Bre93, MSV98, McA93].

Complementing psychoacoustic work on the cognition of rhythm come issues of the best logical representations for rhythm in computational work [Dan93, Hon01, Mar00]. These studies may themselves give insight into the information theoretic scope of mental processing. A team at CNMAT (Berkeley) propose a cross-culturally applicable representation for rhythm in computer music based on Bilmes' notion of the tatum [IBWW97]. From the work of Jeff Bilmes [Bil93] and Desain and Honing [DH93], the representation of expressive timing is no longer couched exclusively in terms of a master tempo curve, but would follows a combination of low frequency tempo variation and high frequency timing deviation; Desain and Honing also criticise the cognitive basis of tempo curve perception, arguing that it is an abstraction from the reality of discrete observed events.

Good reviews on the perception of rhythm are [Cla99, Fra82, Han89]; the excellent Parncutt article [Par94] integrates much work. The act of musical performance and expressive timing is covered by [Gab99, Pal97, Slo82].

[4]Jitter in timing production is due to noise in neural processing and in mechanical production (musculature etc); message passing through biological systems is inherently noisy since the same starting condition can never be repeated exactly [Mic00, page 86].

2.2 Pertinent Scientific Results for Beat Induction

The perceptual present is an important factor in integrating evidence of pulsation. Summarising research, Pöppel cites a three second temporal integration limit for the 'now', with a 300 mS minimum for separable conscious events [PW99]. Parncutt [Par94, page 437] adopts the 200-1800mS range of trackable pulsation levels, corresponding to a 33-300bpm range of musical tempi. He further notes the echoic store for registral memory extends 0.5 to 2 seconds (page 428) and that the maximum number of distinct events that can be taken into consideration in determining the rhythmic context is twenty-five (page 451). Mates et al. write that 'only if successive stimuli fall within one integration period, can motor programs be initiated properly' [MRMP94], the maximum for the window being three seconds. Experimental data recording the anticipation (most consistently at preferential tempi) or late reaction (outside the integrating window) to events is provided in figures in the source. An fMRI study showing the comparison of activations for 0.6 second and 3 seconds intervals is [LM03].

In performance, fast note sequences are dealt with through chunking, using motor sequencing- 'subdivisions of the beat (i.e., individual notes) are not directly timed, but are produced by overlearned motor procedures that specify movement patterns that have as their consequence a definite timing profile' [Cla99, page 495].

In a spontaneous tapping study to test hierarchic perception and referent tapping rate, Drake et al. [DPB00] demonstrated that musicians have the greatest range of available hierarchical levels and select the slowest pulses as the tracking level. This is linked to dynamic attending theory and developmental experiments in [DJB00][5]. In dynamic attending the time course of event streams are tracked by focused concentration on the most likely predicted locations, perhaps akin to the careful allocation of information processing resources [JY93].

Whilst many have followed Fraisse in centering the preferred tempo curve at 600mS (100bpm), a paper by van Noorden and Moerlants [vNM99] revises the peak to 500-550mS in the light of a survey of tempi across (Western) musical styles, a refreshing study of preferred pulsation rate against polyphonic audio, and a model of resonance applied to experimental results from three previous studies.

Much useful data on human tapping tasks is provided by Repp [Rep01], who also finds a detection threshold for tempo changes of 2%[6] and that acceleration is more sensitively spotted than deceleration. Period correction is fast and under conscious control, whereas phase correction is slow and the approach taken for subliminal tempo changes. Synchronisation is fine tuned outside of conscious attention. A recent article on the neurobiology of entrainment with respect to cerebellum damaged patients [MLM$^+$03] found evidence to suggest that beat induction processes exist in a low level unconscious form early on in the neural auditory system (pre-cerebellum) and in separate high level attentional processing distributed perhaps in the basal ganglia and cerebellum.

The human ability to follow and then anticipate musical events is astounding. Fraisse notes that 'what is important is not the regularity but the anticipation' [Fra82, page 154]. He reveals that synchronisation can occur from the third heard sound, can track complex rhythms and not just isoperiodicities, and is also maintained to an accelerating or decelerating sequence, though the effectiveness is inversely proportional to the gradient. A hypothesis of tracking is tenaciously maintained from early evidence:

> 'The first perceived pattern tends to impose its structure on the later patterns ... this fact confirms the importance of predictability as the basis of rhythmic perception' [Fra82, page 162].

Musical perception is not exempt from higher level categorisation effects- 'even trained subjects could differentiate only two, or at most, three durations in the range of perceived durations (below two seconds). If the durations were more numerous, confusion arose.' [Fra82, page 168]. This is a sobering thought, though trained musicians surely deal regularly with many more ratios: what status for tuplets? Further, listeners will systematically overestimate short durations and underestimate long durations [Cla99, page 475]. Dotted rhythms are produced in a tempo dependent manner: 'the ratio long time–short time is maximal at the spontaneous tempo' [Fra82, page 168]. When moving away from this rate to the limits of production, there is no longer any great distinction. The common deviation from the exact scored durations for such figures in musical practise is noted by Weisberg [Wei93]. Parncutt [Par94, page 444-5] subsumes the dotted quaver and swing (the notes inégales (unequal)) in

[5] Informed selection of the tactus solves Temperley's concerns about having the fastest occuring level, being the greatest common divisor, as the referent [Tem01, page53]: the tactus should be placed intermediate within the preferred tempo range [DCP99, page 192]

[6] This varies significantly between subjects, and Repp previously placed discrimination at 4% in an earlier review [Rep00b, page 129].

his salience model as providing indication of the wrapping beat location, but not of any subdivision on the order of the events themselves. Context is all important in classification. Clarke makes an example of the IOI sequence [600, 400, 1000] which in a duple metre may be interpreted as 1:1:2, in triple as 2:1:3 and in quintuple as 3:2:5 [Cla99, page 490]. Finally, perhaps important in the modeling of free jazz improvisation is another result due to Fraisse. In arrhythmia higher ratios are less frequent, and a ratio of 1:1 predominates in all production [Fra82, page 165].

Chapter 3

Beat Induction

3.1 A Review of Beat Induction Models

There have been many attempts to produce tempo tracking and beat induction models, none entirely successful; admittedly no human listener is an expert in the rhythmic intricacies of all the world's musical styles. Whilst most models are broadly inspired by human audition, for real-time application, where full computational auditory models are too slow, pragmatic methods and optimisation shortcuts are often invoked. On the other hand, any engineering success in this field may contribute to an understanding of the possibilities of auditory signal processing.

The essential differentiators of models are whether they are causal (do not require the entire input to operate), real-time (calculate within the response constraints of a perceptual present) and act upon a symbolic representation or pure audio input. A number of models, particularly from the cognitive science sector, assume a clear symbolic starting point, with isolated and marked onsets, though there are also engineering approaches [CK03] that start from an equivalent MIDI representation of events.[1]

The focus of this study is slanted towards working with audio input, causal, real-time models, the ultimate authority being an experienced human improvisor.

Most onset detectors work in a way that loosely follows the human hearing model. The incoming audio signal[2] is split into some set of sub-bands (or a set of fixed filters over the most sensitive parts of the human hearing range), and for each a form of temporal integration of energy is applied (using convolution or some filter like a leaky integrator).[3] Derivatives of these signals may be taken rather than the pure values. Downsampling may be used for lower frequency bands, and on the smoothed envelopes to reduce the information processing requirements. A second stage copes with the selection of peaks in the smoothed envelopes for signal energy in each band, by some absolute or adaptive thresholding, and by considering the combination of results across sub-bands. Scheirer notes that the combination used by humans is non-trivial 'some sort of cross-band rhythmic integration, not simply summation across frequency bands, is being performed by the auditory system' [Sch98, page 590]. Onsets may correspond to perceived onsets (p-centres [VR81, Mar81]) or by certain measures or corrections to physical onsets (the initial sample of a signal event).

Hainsworth's review [Hai04, pages 46–57] categorises models as rule-based, autocorrelative, oscillating filters, histogramming, multiple agent and probabilistic. Gouyon and Meudic [GM03] allocate computational models by the processing of symbolic or audio data, and further by a top-down (knowledge based) or bottom-up (signal analysis) approach. The historical trend is from purely symbolic manipulation towards more recent audio based models, with the eventual goal being a causal real-time system working straight from an audio input. There is a definite recent trend towards Bayesian inference models in statistical signal processing, alongside an explosion in interest in temporal structures of music related to content description research, seeking salient features of signals as audio descriptors for

[1] The independence of onset detection from beat induction is a reductionist convenience. It is entirely possible that event detection and metrical induction are not cleanly separable in this manner, especially for complex polyphonic signals-witness the continuing relative success of Scheirer's entirely signal based approach.

[2] A stereo signal is usually mixed to mono before submission to an onset detector. A two channel comparison may be interesting, however, due to interaural intensity and phase differences, and filtering (very significant for auditory scene segmentation). Information must be reconciled when peak picking with differences in masking and energy integration in the two ears.

[3] Alternatively (though closely related in signal processing terms) a frequency domain transform is applied via FFT or wavelet transform, and features sought over frames from an examination of changing phase vocoder information (phase and amplitude of FFT bins).

classification purposes. The review undertaken here details the main published models and available technologies for beat induction, with an emphasis on recent attempts and the best models as starting points for implementations.

Assessing and comparing beat induction models is an issue in itself, and an increasing concern of more recent reports [Dix01a, Dix01b, GM97, Hai04, Kla03].

3.1.1 Rule-based Symbolic Models

The earliest metrical induction models were formed of a small set of heuristics, operating on a quantised (cleaned up or noiseless) symbolic input of the inter onset intervals. Lee [Lee91, page 64] summarises and compares a number of models by Steedman, Lee, Longuet-Higgins, Lerdahl and Jackendoff, and Povel and Essens. Five of the models (those from the early eighties) are compared with novel experimental results in his article, and a revised model built. Lee and Longuet-Higgins' models are causal and Lee shows a refreshing concern about processing considerations: 'a general problem with the proposals ... is that they take no account of the fact that the listener has to build up a metrical interpretation in the course of listening to a sequence [Lee91, page 73] .

It should be noted that whilst Lerdahl and Jackendoff's music theory [LJ83] has well-definition and preference rules for metre, those authors provide no explicit algorithm. They have been computationally applied by Temperley using dynamic programming techniques [Tem01].

The Povel and Essens model [PE85] argues for the existence of a hierarchical internal clock as the basis of beat perception; their mathematical treatment is simply of a single isoperiodic unit, determined by exhaustive search through all possibilities of duration less than half the input sequence length. Three rules for accent based on proximity of events mark up input sequences, and the best clock is chosen as the least negative option from a weighting function. They do confess the existence of input patterns that do not admit a clock solution, but claim a musical significance for clocks in an experiment which gave subjects the same target patterns as the model:

> 'the prediction that subjects would have more difficulty in forming an accurate representation of a sequence, the weaker that sequence induces a clock, is supported by the data' [PE85, page 425].

Their model is not causal, requiring the whole pattern for assessment.

Eck [Eck00] provides a comparison of the Povel and Essens model with two by McAuley and Semple and one of his own (TNPOS), using experimental data from the two earlier studies and Parncutt's salience work [Par94]. He thus distinguishes models that work by positive evidence, and those functioning by negative, and in TNPOS uses normalisation by the clock rate to avoid a problem whereby the fastest pulsation rate always won. TNPOS is shown to be a relatively effective model, even though it has but one degree of freedom compared to the four of the Parncutt.

A rhythm parser by Rosenthal [Ros92] must operate on the whole phrase since it parses both forwards and backwards. MIDI data is the starting point, though the assumption seems to be that expressive timing and errors are filtered out without trouble by an unspecified quantisation program. Rosenthal provides an intriguing excuse (provided by Caroline Palmer) for his model's avoidance of velocity information; pianists do not produce accents on structural downbeats with any consistency, and he has already factored out timing deviations that might mark such salient points. The rule system is not that clearly presented but seems based on updating an IOI hypothesis within a shifting temporal window; like most of these models, it will not cope with tuplets.

An innovative review and extension of the rule based systems of three models by Longuet-Higgins and Lee is presented by Desain and Honing [DH99], who test these theories on large databases of ametrical and metrical material, probing for effective rule combinations and optimising free parameters. They acheive an 80% success rate with optimisation of an as yet unpublished Longuet-Higgins model. An intriguing comment at the end of the paper notes the feasibility of a research project based on genetic programming optimisation of beat induction rules.

Whilst some of these models are causal, the fact that they require a quantised input means that they will only be effective as a second stage to a tempo tracking duration categoriser. Many of the models are also discussed by Clarke [Cla99, pages 482-9].

3.1.2 Signal Processing Models

Barry Vercoe's early research into audio based beat induction [Ver97] eventually found its way into Csound's spectral opcodes [Ver00]. Vercoe is inspired by auditory modeling, using a constant Q transform with 96 bands (which calculates one hundred frames per second) and weighting based on Fletcher-Munson curves. The spectral difference between frames becomes the onset signal. This is passed through basic recursive filters for temporal integration, summed, and assessed by a modified autocorrelation function which can find phase as well as period information, though the details are in the source code but not the book chapter.

Also in the MIT Machine Listening group, Eric Scheirer's celebrated 1998 article [Sch98] introduced the use of comb filters for tempo hypotheses. Each sub-band's amplitude envelope signal is passed through a selection of comb resonators; these will have noticeable outputs when pulsations match the rate or simple rational multiples and divisors of the rate (superior to correlation). Phase can be determining by considering how the model would predict future events if the audio input suddenly became silent, that is, by looking in the comb filter delay lines for peaks. The drawback of the model is in providing enough comb filters to cover the tempo space, and in following a continuously changing tempo. Scheirer's code implementation is somewhat inefficient, using a time domain rather than FFT convolution for the 200mS half-Hanning integrator[4], but eminently adaptable to a real-time and causal model. Note that the time delay due to convolution calculations can be caught up by post-rationalisation of the beat location, as occurs for temporal integration in human audition.[5]

Masataka Goto [GM99, Got01, GM97] has a beat induction model specialising in 4/4 popular music with or without drums, which uses an FFT frame metric based onset detector, a chord change detector working from provisional beat hypotheses, and a drum pattern detector focusing on low frequency kick and broadband snare sounds. A combination of heuristics and correlation techniques allows a 'hierarchical beat structure' of quarter-note, half-note and measure to be built up. Multiple interpretations are considered by multiple agents. Full implementation details are not disclosed by Goto.

Anssi Klapuri's advanced model [Kla03] combines and generalises work by Scheirer and Goto. In particular, Klapuri introduces a difference of logs (ratio) instead of a first order difference function for spotting onsets. Bayesian probabilistic methods are used to determine the best metrical hypothesis, with constraints on continuity over time. The estimation of metre seeks three pulse levels, the measure, beat and *tatum* (temporal atom or time quantum). The tatum is also known as the clock or tick in other work, and corresponds to the fastest induced pulse, as approximately the greatest common divisor of the IOIs; the term was introduced by Jeff Bilmes in honour of Art Tatum [Bil93]. Interestingly, the Large Kolen model [LK94] and the Scheirer comb resonators are compared as beat induction stages, and come out performing equivalently under optimisation of parameters (Klapuri chooses the Scheirer for model simplicity). In common with all other beat induction models, there are certain rather arbitrary tweaks and parameter choices within the model (witness the phase estimation weighting in section 2.2.1), which shows that there is much work on optimisation and comparison to do, and much scope for exploration. The final part of the paper assesses Klapuri's model against those of Dixon and Scheirer (the two main models with publicly released source code), on a large database of 478 pieces, with the Klapuri model coming out best.

Jarno Seppänen's impressive master's thesis [Sep01] was undertaken in the same lab as Klapuri. Onset detection is followed by a combination of Parncutt's phenomenal accent model and a tatum grid estimation to estimate the beat by a final Bayesian stage. A novel error function is presented for finding the best fitting greatest common divisor (GCD) to an IOI histogram. Seppänen also considers the optimisation of the acoustic feature set for accentuation.

Iroro Orife [dO01] takes a polyphonic audio signal and separates out components by Independent Subspace Analysis (ISA) [FCL02]. Each line so revealed is passed through an onset detector modelled on Scheirer and Klapuri, and then the tatum induced using Seppänen's GCD error function. Having the tatum for a number of parts in a polyrhythmic drum pattern allows a useful decomposition of complex rhythms.

Simon Dixon has developed a beat tracking model, and compared it to other researcher's work [DPW03, Dix01a, Dix01b]. His audio analysis is remarkably straight forward, simply utilising a highpass filter, maximal amplitude in some short-time window for enveloping, and a four point regression to curve fit the slope for a thresholded onset picker. An IOI histogram method finds likely tempi, then period and phase hypotheses are scored by multiple agents, which are created and destroyed over analysis time based on new hypotheses (assuming that new onsets are possible new zero phase points) and redundant or old data.

[4]The magic number 200 mS appears in human energy integration studies; masking effects do not occur outside a 200mS temporal window [vNM99].

[5]An immediate adaptation of Scheirer's model that springs to mind is that of allowing attention to change the delay length or number of the comb filters, so that they focus onto the best fitting tempi- this would make the system similar to multi-agent hypothesis systems by Goto, Dixon or Large.

The Music Technology Group at Universitat Pompeu Fabra has many relevant publications in this field. Fabien Gouyon [GHC02, GM03, GFB03, GH03a, GH03b] is the main co-author, often working from the angle of audio content analysis research, and has published a number of models for beat induction. After a maximum amplitude per frame integrator for onset detection, beat induction is accomplished by IOI histogramming, seeking the tatum [GHC02]. In a later paper [GH03a] autocorrelation is used to find significant periodicities amongst audio features taken in analysis frames, these being matched to tatum positions to find the beat period and phase. An interesting application presented by the group is the modification of swing in recordings [GFB03]. Gouyon and Herrera [GH03b] show a concern with finding signal features that correspond to downbeat accentuations (beat descriptors), marking out the temporal centroid as one very useful quantity.

The Centre for Digital Music in Queen Mary University of London is publishing an increasing amount on automatic transcription and related digital audio processing. They provide computationally tractable real-time models for transient detection and onset detection [DDS01, DDS02, DBDS03]. Algorithms use instantaneous frequency for the transient test, with a multiresolution separation via an octave spaced filterbank and adaptive thresholding for automated rather than manually set detection levels. The later onset detector [DDS02] makes use of further measures based on signal energy but uses additive differences rather than log ratios. A mathematical treatment of the problem of phase and amplitude information in detection leads to a complex plane formulation of great clarity and applicability [DBDS03]. The transient versus steady state detector [DDS01] has processing applications outside of onset detection.

Aside from Klapuri, a number of other researchers are currently investigating Bayesian statistical models of rhythmic inference. The Bayesian approach sees the calculation of the maximum a posteriori (MAP) state based on prior knowledge and a estimated likelihood [Mac03]. The calculation of the MAP may require the advanced mathematical tools of Monte Carlo integration and optimisation. Exponents of this methodology include Cemgil and Kappen [CKDH00, CK03] and the aforementioned Raphael [Rap01a], who provide models of rhythm transcription from symbolic data. Steve Hainsworth's recent PhD [Hai04] describes the most complex inference model to date using sequential Monte Carlo[6], with an audio signal onset detector based on FFT methods. The metric used to compare FFT frames to discover an event resembles Klapuri's log function; Hainsworth and Macleod [HM03] compare possible detection functions. Hainsworth assesses his causal but non-real-time beat inducer on a large hand marked database, comparing his model to the Klapuri, with the Klapuri just edging out in front.

There are plenty of further miscellaneous attempts and peripheral studies; some recent articles are highlighted here. Smith's original approach [Smi96] is a demonstration of the facility of the wavelet transform. Starting from MIDI information, a wavelet transform is taken at a sampling rate of 200Hz on impulse signals representing the onsets. Scalogram and phasogram features reveal expressive timing and periodicities. The periodicity transform discussed in [SS01] is only applicable to stable tempo signals, does not determine phase, and seems to assume foreknowledge of the tempo in the choice of sampling rate! Another non-standard rhythmic representation is that of Foote and Uchihashi's [FU01] beat spectrum, which measures novelty based on an autocorrelation matrix acting over the bins of FFT frames (their interest is in audio content categorisation). Also for music classification and retrieval, Tzanetakis et al. [TEC02] define the beat histogram, derived from a three second wavelet transform, envelope extraction on the octave frequency bands and autocorrelation. Uhle and Herre [UH03] combine a number of methods from prior papers and again use autocorrelation; they estimate tatum and periodicity before bar length and time signature. Their paper does have a refreshingly non Western test database and perspective. Jensen and Andersen [JA03] test six possible features as markers of onsets; high frequency content for an FFT block size of 2048 comes out best. An IOI histogramming method is used for beat induction. Temperley [Tem01, pages 39-42] also describes an IOI histogram method for MIDI data with a weighting by the square root of the beat interval for older events.

3.1.3 Cognitive Models

Many models of the perception and production of rhythm have been devised in cognitive science to explain experimental findings, and may be adaptable to practical tracking systems. The act of entraining to rhythm can be modeled as the adjustment of the phase and period of a self-sustained oscillator due to weak interaction with some external force or another appropriate oscillator [PRK01, page 8].

The earliest historical models are timekeepers, based on control system engineering for motor production [BPD00, VW96]. They are discrete, linear and entirely based on IOIs. Their application is typically to explain variances in timing data in the production of isochronous sequences by continuation or by

[6] Also known as particle filtering- each particle can be thought of as one beat hypothesis or individual agent of a probabilistic multiagent architecture.

sychronisation, hence they are somewhat divorced from the context of musical rhythm. Yet the power of even a simple linear model to explain experimental results is demonstrated by Repp [Rep01]. In comparison, non-linear oscillators are usually continuous dynamical equations which need to be discretised for computational use.

Edward Large has developed a series of oscillator models to explicate entrainment to rhythm [LK94, LJ99, Lar00, LP02], in collaboration with some eminent psychologists. The most widely disseminated is the earliest, a model [LK94] based on the sine circle map [PRK01, pages 201-10]. An oscillator is described by non-linear difference equations for updating period and phase given a sequence of IOIs. In order to track multiple hypotheses, independent entrainment oscillators are run with favoured rates covering the tempo space. This work is adapted by Toiviainen for his jazz automated accompanist [Toi98, Toi01]. Toiviainen notes some errors in the original paper, and corrects the model for real-time use. Large's later revised models see him introduce Mari Riess Jones' theory of attentional dynamics [LJ99], through an adaptive focusing process that updates oscillator states selectively based on how well new evidence (a new IOI) fits in with an oscillator's current assumptions. The Large models have been extensively tested against experimental evidence [Lar00, LP02], and the mathematics tweaked and updated. Large has also placed an onset detection frontend before the oscillators [Lar00, page 547], based on a hearing model from Slaney's auditory toolbox for MATLAB. This is strange in the context, given that he starts from MIDI data, and synthesises it into audio.

Connectionist models, or neural nets, utilise a sub-symbolic representation of neural activity based around abstracted neuron components. A strong body of supporting theory discusses their abilities to perform categorisation and non-linear function approximation, and shows the intimate links with Bayesian inference [Bis95, Mac03]. Musical and signal processing applications are discussed in [TL91, GT99, LU97]. Desain and Honing's connectionist quantiser [DH92] is a rhythmic categoriser based on a simple net. As Clarke notes [Cla99, page 487], metre is an emergent property here of the principle of seeking a simple integer ratio interpretation of the evidence. A Fitzhugh-Nagumo neuron (which has dynamics close to a real spiking neuron) is the basis of Douglas Eck's beat induction attempts [Eck01, Eck02]. He demonstrates some potential for the model, though the synchronisation is hardly immediate and the model is weakly equipped to handle a performer's expressive timing deviations. The recurrent timing nets introduced by Cariani [Car01] are a neurologically plausible relative of autocorrelation techniques. By a sufficiently large set of feedback delay lines, possible correlation lags are covered and a causal detector of activity at different rates established.

To return to the sphere of timekeeper models, the dynamics of motion have been explicitly linked to the perception of beat, not just the production, by a hypothesis of a motor component in auditory tracking tasks. Neil Todd [TOL99, TLO02] approaches beat induction in terms of a simulation of the physiological and neurological process. He involves a musculoskeletal control system which, even in the absence of actual motion, is a necessary component in the loop for beat tracking. A full auditory model processes a signal eventually into an auditory image (periodicity against cochlear position over time), which is analysed by three dimensional filters. Some parts of Todd's work court controversy: the involvement of the vestibular system in hearing, the denial of any internal clock, and the existence of 3-D spatiotemporal filters in the auditory cortex. Unfortunately, in terms of implementation Todd's model is not practical for real-time use, though it is of course causal. A mass-spring robotic arm was introduced to Eck's beat induction model to simulate preferred tempi in [EGP00]. The authors assert that a physical motor component is needed in beat induction models because an 'interaction between body and brain in developing infants cannot be ruled out as an important factor in beat induction learning'. Large has also attempted to justify his oscillator model in terms of human motion [Lar00].

3.1.4 Commercial Systems

Commercial systems are primarily oriented towards bedroom dance music production though score based accompaniment systems do now exist as a commercial reality (http://www.smartmusic.com). It is likely that during the course of this thesis new beat induction systems will come onto the market. With no publications or access to the source code, commercial systems are of course hard to assess and impossible to adapt to compositional needs.

A number of commercial devices, such as the Korg Kaoss Pad, have a 'tap tempo' facility for establishing the period and phase in enough accuracy to sync up fx delay lines and loop sampling. *Beat matching* is the fundamental operation that good DJs master in order to synchronise an outro of one track with an intro of a second for a crossfade that maintains the continuity of the clubbing tactus. Traktor (www.native-instruments.de) provides two virtual soundfile decks with DJ facilities like tempo and pitch adjustment (with tap tempo), looping, cueing and EQ. Non-realtime analysis when importing soundfiles can establish a beatmap which allows the estimation of pulse and synchronisation between decks. Real synchronisation requires knowledge of beat in the two sources to be matched because phase errors will accumulate over time whenever there is a period error, and tap tempo determination of tempo is not that accurate. Since Reaktor only operates on existing soundfiles and not on live input it could make use of faster than realtime analysis to look ahead and predict future beats without any causal algorithm restriction.

In the vein of Traktor's beat maps, commercial jukebox solutions exist, where semi-automated markup of tracks in a database is the technique for beat matching. There is an available technical report by Cliff on this process [Cli00].

Ableton Live (http://www.ableton.com) cannot automatically beat match to live input but does produce beat maps on the fly on audio files imported while the sequencer runs. The time warping of onset points allows sync between running loops. Many commercial VJ softwares also have manual tapping based synchronisation, or use audio analysis, to allow for a 1-1 correlation between audio and visual events.

Recycle (http://www.propellerheads.se/) is a tool for the automatic segmenting of soundfiles, principally drum loops. Onset detection based slicing allows recomposition from the component sounds, including the quantisation of the timing data with different degrees of swing. Other segmented sounds may be manipulated to match a given timing template. Similar segment and retime facilities exist in a number of other studio softwares including the Beat Inspector in Pro Tools, and Logic Audio's Groove machine. The Groove Control by Spectrasonics is a powerful version of this technique, though restricted to the pre-analysed loop library it comes supplied with. Gouyon assesses these softwares [GFB03] finding them lacking in sound quality (for stretching) and flexibility of beat induction compared to engineering methods still in the research domain.

3.1.5 Available Implementations

The Dixon, Scheirer and Vercoe public Csound code is available, and also the MTG CLAM library is GNU GPLed. (http://www.iua.upf.es/mtg/clam/) and may contain some useful source code.

A number of models have been implemented in Cycling74's proprietary Max/MSP and for the freeware PD graphical programming languages for audio. Miller Puckette has provided beat induction and pitch tracking objects (bonk and fiddle respectively) and detailed them in a 1998 ICMC paper [PAZ98].

Seppänen has worked on some PD objects (http://iem.kug.ac.at/mailinglists/pd-list/2002-09/010502.html). The Large Kolen and Toiviainen oscillator models have been implemented by Olaf Matthes for PD (http://www.akustische-kunst.org/puredata/maxlib/). Other researchers are converting some of the papers above to performance implementations. Ari Lazier is coding the Klapuri paper as a beat external (http://soundlab.cs.princeton.edu/software/beat.html). Tristan Jehan is working on the Scheirer model (http://web.media.mit.edu/ tristan/). Neither has released any source or product yet.

SuperCollider currently lacks any beat induction models except for the functionality I myself have started to build, detailed below.

3.2 A Critique of Beat Induction

It should be productive to consider cases where inference of metrical structure is difficult for human observers. Comments earlier about the perceptual relevance of tracking tatum, tactus or timeline are important here[7]. Even within a genre, composers and performers will play with idioms, aiming for a novel meeting and teasing of expectancies. The rules of musical styles may only be learnt by extensive training on examples, and be specific to those databases. Conventions of style will cheerfully mislead naive tracking programs. Reggae and ska put a lot of emphasis on offbeats, yet chords change on the bar. Recognition of style should inform a beat inducer how to interpret evidence gleaned from both chord change detection and onset detection. There is no guarantee of a universal tracking algorithm.

In fact, there are many reasons that beat induction can be a badly posed problem. The beat often admits multiple solutions: 'a central assumption...is the inherent ambiguity of the underlying pulse (tactus) and meter of a rhythm' [Par94, page 423]; GTTM's musical *preference* rules underscore this point. 'We have seen that bars, even those of a classical musician such as Bach, and even when played by the same orchestra, can be perceived in multiple ways' [Fra82, page 175]. Solutions in transcription may be accurate only to homomorphism, since a composer can decide the notated scale of the beat (as

[7] How should models cope with a 5/8 time signature? With separate cases for 2+3 or 3+2? By treating an eighth note tatum, measure length pulse or timeline from the additive structure?

2/2 versus 2/4, say). Povel and Essens point out that the IOI pattern 31111213 admits beat solutions of size 4 and of size 3 [PE85, page 432]. Expressive timing comes to the rescue– this figure would surely be resolved by expressive accentuation if this were to be performed. Natural music making is less ambiguous than bare pathological IOI examples: 'meaningful musical material does contain many redundant cues to the metre' [DH99, page 37]. Ambiguities still exist in metrical levels, particularly for less trained observers; the most common error in a beat induction algorithm is finding the wrong metrical level for the beat when many are possible (untrained humans also make this mistake). Toiviainen and Synder explicitly discuss the switching of attention between metrical levels [TS00].

Scheirer has always been set against the grand design of perfect score reproduction, and his point is easy to see especially for contemporary scores.[8] Not all gestures can be communicated, some are only in the mind of the performer. At the very least, general beat induction models would require some known structures existing as a template, like the tāl or a jazz standard chord sequence which maintain an improvisation framework.

Instrument (timbral stream) recognition is not part of any polyphonic audio beat induction model save for the Orife, but may provide important cues for stylistic context essential to tracking.[9] The MIT lab has a stance [MSV98] on the use of auditory scene analysis rather than pitch models founded in music theory. Scheirer [Sch99] argues that in human transcription, not every component note is heard; instead, chords are heard as unified entities. Tracking software may operate best at a feature extraction level based on auditory scene analysis, rather than pursue a false goal of pure separation of all events. Scheirer, Dixon and Goto [Dix01b, Sch98, Got01] all note that beat induction does not require full score knowledge to operate- the average listener knows little music theory, and yet they tap along happily to the beat.

The dynamics of disruption of a listener's frame of interpretation may play an interesting part. There is an unwillingness to change metrical hypotheses in the face of continued syncopation. Heuristics in the rule-based systems tend to take the first established beat as the most reasonable. A beat induction algorithm would be better informed if it knew when to best abandon a cherished hypothesis.

Swung and dotted rhythms are bound to perturb measures of the IOI histogram and tatum. Can atoms ever be simple timelines themselves? Parncutt [Par94, pages 444-5] finds that the evidence suggests not; swing will only cause a sensation of the enclosing beat.

This thesis is assisted by some pragmatic points. It is not necessary to produce a general beat induction model, just one specific to some interesting compositional application. Full metrical knowledge is not required; if you can match the beat well, you can phase shift quite trivially to get to the measure and hyper-measure sync (assuming fixed and known time signature). A system can be given initialisation clues (by a manual tap tempo process) before it tracks expressive timing as best it can.

There are possible issues with onset detection errors and delays, though a robust model should cope. Temporal integration on an incoming signal has a wait time of FFT frame size or filter delay; for instance, Vercoe's constant Q transform imposes a wait for the lowest frequency sine. The perceptual centre does not correspond to physical onset. These latencies (and any audio output delay) can be compensated as soon as a predictive model is in place, for incoming events can always be post-rationalised and future action then scheduled to synchronise.

For the most difficult case, of real-time tracking of a human improvisor without score or rehearsal, it is possible that a solution is unobtainable. Human improvisors can track each other given sufficient commonality of assumptions or if a continuous shadowing is undesired, and will do so with a graceful degradation of performance (i.e. good musicianship). Dixon distinguishes predictive and descriptive beat tracking[Dix01a, page 51]; whilst musical structure might be described after the event, fully predictive tracking of expressive tempo modulation in novel music may be an impossibility. Experiments on extreme on-the-fly tracking between human musicians could help to resolve this.

[8] A complex progression of time signature and tempo changes, say, Boulez's *le marteau sans maître*, is not likely to give rise to a perceptual resultant accurate to its every twist. That this score is necessary for the unfolding of the music for the performers themselves is still quite possible.

[9] Whilst an aggregate of cues in polyphonic music may be helpful, beat tracking of monophonic instruments is still possible.

Chapter 4

Research Schedule

4.1 A Performance System: *bbinduct*

Figure 4.1 shows a basic signal flow diagram for a performance system under development for live interactive computer music. The majority of the system exists already as a product of previous research, and is based around the SuperCollider Server [McC02] computer music language for real-time synthesis and processing. The live coding interface and mixer [CMRW03, Col03] allows the use of the SuperCollider interpreted programming language to edit the code specification and running functionality of audio processes on the fly. The algorithmic composition system is *bbcut* [Col02], an established tool for live generative audio cutting. OSC (Open Sound Control [WF97]) is a network protocol for audio control messaging, and is output by running bbcut instances to pass their state to an external application for video. This technology allows synchronised video processing, and is used in a live implementation with a VJ [CO03].

The current capabilities of bbcut include a real-time causal onset detector, shown in figure 4.2, but no automated beat inductor. The performance system has tempo controls for beat matching, with manual tempo set, tap-tempo, phase jog and some experimental algorithms for gradual period and phase transition.

The concern of this thesis is in the top half of figure 4.1, the extraction of features and beat induction on an incoming audio stream, and how it may improve interaction with human musicians, or sound processing possibilities. The target audio stream is imagined to be that of a musician on a monophonic instrument, or a polyphonic audio source, perhaps an ensemble or disk streamed audio. A likely target is that of the human voice; in fact, hip hop culture's beat boxing is a contemporary vocal technique that may provide interesting control possibilities.[1] Whilst audio input is the general and flexible case I wish to treat, control streams (MIDI, OSC, USB) from electronic interfaces are a possibility as additional inputs. Video tracking will not be considered as a source of tempo information, though there have been studies which allow dancers to become conductors [Gue03].

The tasks may be enumerated:

1. To track the pulse of polyphonic audio recordings, initially for percussive and metronomic dance tracks, secondarily for expressively timed tracks. Assumptions of relatively stable tempo and 4/4 should make this an accessible goal through adaptations of existing beat induction models.

2. To track a monophonic or solo instrument from a microphone, extracting salient rhythmic features (periodicities, patterns like timelines, metrical structure etc) for computational use; these to act as a direct conductor by inference of the pulse or as a more abstract contribution to generated music.

3. To investigate further useful features (not necessarily rhythmic) for real-time interaction and control from a human instrumentalist.

4. To apply these routines as the frontend to an interactive system with compositional output making use of bbcut, with any necessary modifications to that system to integrate the audio tracking.

If a beat induction model is very CPU intensive it will not run on the same computer as the composition software. However, OSC would allow a two computer system to be setup, the beat induction

[1] Jazz scat singing as a sound source was independently suggested by Jouni Paulus [PK03]

Figure 4.1: Components and information flow in the proposed interactive system

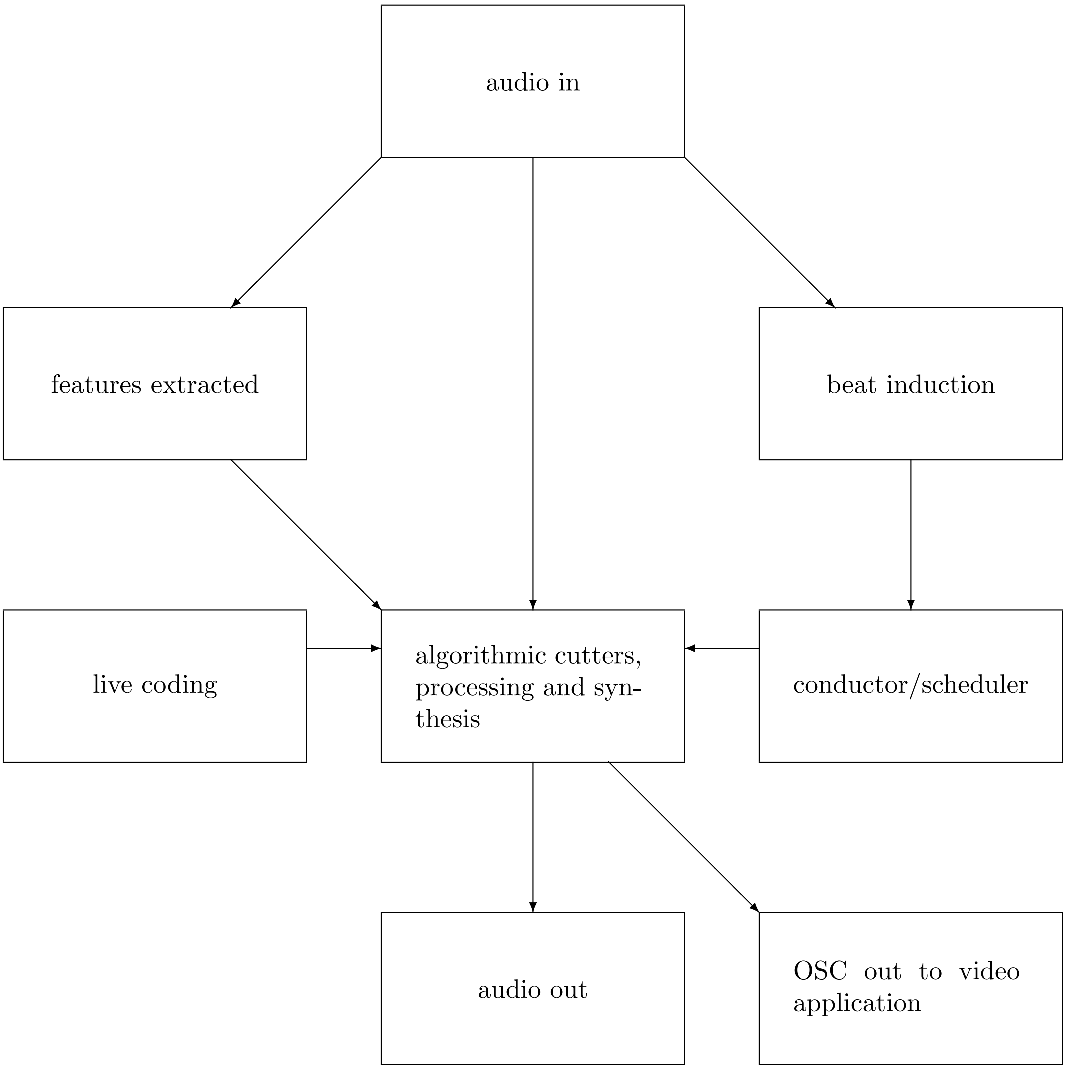
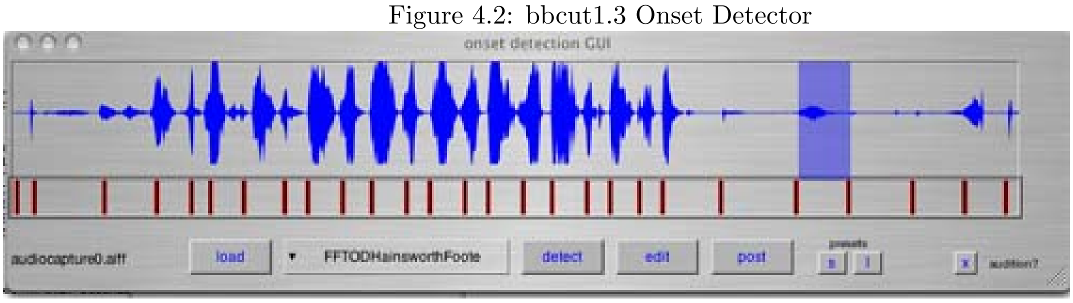


Figure 4.2: bbcut1.3 Onset Detector



computer passing through tempo and phase information to the live composition computer. It is trivial to send the audio in to both machines. There is a possibility that the existence of a freeware beat induction object by a third party (as mentioned for PD above in section 3.1.5) would allow the running of an off the shelf solution; beat can be passed as OSC messages to SuperCollider. This is unlikely to provide the necessary flexibility, and will have to be assessed.

For beat induction itself, the most promising model to follow is the Klapuri work, perhaps utilising the Seppänen adaptation of the Parncutt pulse salience model. All the main beat induction papers can provide useful methods, particularly the Goto, Scheirer and QMUL papers.

I have already implemented Jensen and Andersen 2003 [JA03] and Hainsworth 2004 [Hai04] onset detectors of straight forward design, though without complex peak picking methods. These were used to build onset detection UGens in C as SuperCollider plug-ins sufficient for the latest bbcut release. I have private (unreleased) Large and Kolen 1994 [LK94] and Toiviainen 1998 [Toi98] oscillator models built in the SuperCollider language itself, ready to connect to live onset detection.

It is very likely that in order to accommodate the sorts of live control of beat and rhythmic material suggested in this project, the bbcut algorithmic composition software, and particularly its scheduling mechanisms, will need revision. This may extend to the scheduling assumptions of SuperCollider itself- the TempoClock object has no immediate phase jump control. Theories of rhythm also suggest new compositional ideas for the automated cutters.

Full assessment and optimisation of the beat induction requires human tapping data across many tempo changes (abrupt and continuous), and a database of onsets and beat locations marked in source material. There may well be existing sources, for this author is reluctant to spend a long time reproducing such data. For instance, Hainsworth, Klapuri, Goto and Gouyon all have hand clapping databases. The often used MMM group piano performance data or Dixon's expressive timing data is another option; in short, contacting other researchers would be a helpful step.

4.2 Acknowledgements

This research is supported by AHRB grant 2003/104481. Thanks to Ian Cross, Alan Blackwell, Juan Bello, Steven Hainsworth and Chris Raphael.

Bibliography

[Aga95] V. Kofi Agawu. *African Rhythm: A Northern Ewe Perspective.* Cambridge University Press, Cambridge, 1995.

[Aro89] Simha Arom. Time structure in the music of Central Africa: Periodicity, meter, rhythm and polyrhythmics. *Leonardo*, 22(1):91–9, 1989.

[Bil93] Jeff A. Bilmes. Techniques to foster drum machine expressivity. In *Proc. Int. Computer Music Conference*, 1993.

[Bil02] John A. Biles. GenJam: evolutionary computation gets a gig. In *Conference on Information Technology Curriculum*, Rochester, NY, 2002.

[Bis95] Christopher M. Bishop. *Neural Networks for Pattern Recognition.* Oxford University Press, Oxford, 1995.

[BPD00] Peter J. Beck, C (Lieke) E. Peper, and Andreas Daeffertshofer. Timekeepers versus nonlinear oscillators: how the approaches differ. In Desain and Windsor [DW00], pages 9–34.

[Bre93] Albert S. Bregman. Auditory scene analysis: hearing in complex environments. In McAdams and Bigand [MB93], pages 10–36.

[Car01] Peter Cariani. Temporal codes, timing nets and music perception. *Journal of New Music Research*, 30(2):107–35, 2001.

[Cas97] Lawrence Casserley. Person to... person?: George Lewis. *Resonance*, 6(1), 1997. http://www.l-m-c.org.uk/texts/lewis.html.

[CK03] Ali T. Cemgil and Bert Kappen. Monte Carlo methods for tempo tracking and rhythm quantization. *J. Artifical Intelligence Research*, 18:45–81, 2003.

[CKDH00] Ali T. Cemgil, Bert Kappen, Peter Desain, and Henkjan Honing. On tempo tracking: tempogram representation and Kalman filtering. In *Proc. Int. Computer Music Conference*, pages 325–8, Berlin, September 2000.

[Cla99] Eric F. Clarke. Rhythm and timing in music. In Deutsch [Deu99], pages 473–500.

[Cla00] Martin Clayton. *Time in Indian Music: Rhythm, Metre and Form in North Indian Rag Performance.* Oxford University Press, Oxford, 2000.

[Cli00] Dave Cliff. Hang the DJ: Automatic sequencing and seamless mixing of dance-music tracks. Technical Report HPL-2000-104, HP Labs Technical Reports, 2000. http://web-new.hpl.hp.com/techreports/2000/HPL-2000-104.pdf.

[CMRW03] Nick Collins, Alex McLean, Julian Rohrhuber, and Adrian Ward. Live coding techniques for laptop performance. *Organised Sound*, 8(3):321–29, 2003.

[CO03] Nick Collins and Fredrik Olofsson. A protocol for audiovisual cutting. In *Proc. Int. Computer Music Conference*, pages 99–102, Singapore, 29 September-4 October 2003.

[Col02] Nick Collins. The BBCut Library. In *Proc. Int. Computer Music Conference*, pages 313–6, Goteborg, Sweden, 16-21 September 2002.

[Col03] Nick Collins. Generative music and laptop performance. *Contemporary Music Review*, 22(4):67–79, 2003.

[Coo01] Perry Cook. Principles for designing computer music controllers. In *New Interfaces for Musical Expression (NIME)*, 2001.

[Cro99] Ian Cross. AI and music perception. *AISB Quarterly*, 102:12–25, 1999.

[CRZF03] Sergio Canazza, Antonio Rodà, Patrick Zanon, and Anders Friberg. Expressive director: A system for the real-time control of music performance synthesis. In *Proc. Stockholm Music Acoustics Conf.*, 2003.

[Dan89] Roger Dannenberg. Real-time scheduling and computer accompaniment. In Max V. Mathews and John R. Pierce, editors, *Current Directions in Computer Music Research*, pages 225–61. MIT Press, Camb, MA, 1989.

[Dan93] Roger Dannenberg. Music representation issues, techniques and systems. *Computer Music Journal*, 17(3):20–30, 1993.

[DBDS03] Chris Duxbury, Juan P. Bello, Mike Davies, and Mark Sandler. Complex domain onset detection for musical signals. In *Proc. Digital Audio Effects Workshop (DAFx)*, 2003.

[DCP99] Carolyn Drake, Laeticia Cros, and Amandine Penel. How fast is that music? The relation between physical and perceived tempo. In *Proc. Int. Conf. on Music Perception and Cognition*, Seoul National University, 1999.

[DDS01] Chris Duxbury, Mike Davies, and Mark Sandler. Separation of transient information in musical audio using multiresolution analysis techniques. In *Proc. Digital Audio Effects Workshop (DAFx)*, Limerick, Ireland, 2001.

[DDS02] Chris Duxbury, Mike Davies, and Mark Sandler. A hybrid approach to musical note detection. In *Proc. Digital Audio Effects Workshop (DAFx)*, pages 33–8, Hamburg, 2002.

[Dea03] Roger Dean. *Hyperimprovisation: Computer-Interactive Sound Improvisation.* A-R Editions Inc., Middleton, Wisconsin, 2003.

[Deu82] Diana Deutsch, editor. *The Psychology of Music (1st Edition).* Academic Press, London, 1982.

[Deu99] Diana Deutsch, editor. *The Psychology of Music (2nd Edition).* Academic Press, San Diego, CA, 1999.

[DH92] Peter Desain and Henkjan Honing. *Music, Mind and Machine: Studies in Computer Music, Music Cognition and Artificial Intelligence*. Thesis Publishers, Amsterdam, 1992.

[DH93] Peter Desain and Henkjan Honing. Tempo curves considered harmful. *Contemporary Music Review*, 7(2):123–38, 1993.

[DH94] Peter Desain and Henkjan Honing. Does expressive timing in music performance scale proportionally with tempo? *Psychological Review*, 56:285–92, 1994.

[DH99] Peter Desain and Henkjan Honing. Computational models of beat induction: The rule-based approach. *Journal of New Music Research*, 28(1):29–42, 1999.

[Dix01a] Simon Dixon. Automatic extraction of tempo and beat from expressive performances. *Journal of New Music Research*, 30(1):39–58, 2001.

[Dix01b] Simon Dixon. An empirical comparison of tempo trackers. In *Proc. 8th Brazilian Symposium on Computer Music*, 2001.

[DJB00] Carolyn Drake, Mari Riess Jones, and Clarisse Baruch. The development of rhythmic attending in auditory sequences: attunement, referent period, focal attending. *Cognition*, 77:251–88, 2000.

[dO01] Iroro F. Ọ. Orife. Riddim: A rhythm analysis and decomposition tool based on independent subspace analysis. Master's thesis, Dartmouth College, Hanover, New Hampshire, May 2001.

[DPB00] Carolyn Drake, Amandine Penel, and Emmanuel Bigand. Why musicians tap slower than nonmusicians. In Desain and Windsor [DW00], pages 245–8.

[DPW03] Simon Dixon, Elias Pampalk, and Gerhard Widmer. Classification of dance music by periodicity patterns. In *Proc. Int. Symp. on Music Information Retrieval*, 2003.

[DW00] Peter Desain and Luke Windsor, editors. *Rhythm Perception and Production.* Svets and Zeitlinger, Lisse, the Netherlands, 2000.

[Eck00] Douglas Eck. A positive-evidence model for classifying rhythmical patterns. Technical Report IDSIA-09-00, IDSIA, Lugano, Switzerland, 2000.

[Eck01] Douglas Eck. A network of relaxation oscillators that finds downbeats in rhythms. Technical Report IDSIA-06-01, IDSIA, Lugano, Switzerland, 2001.

[Eck02] Douglas Eck. Real-time musical beat induction with spiking neural networks. Technical Report IDSIA-22-02, IDSIA, Manno, Switzerland, 2002.

[EG97] David A. Eddins and David M. Green. Temporal integration and temporal resolution. In *Hearing* [Moo97a], pages 207–42.

[EGP00] Douglas Eck, Michael Gasser, and Robert Port. Dynamics and embodiment in beat induction. In Desain and Windsor [DW00], pages 157–70.

[FCL02] Derry FitzGerald, Eugene Coyle, and Bob Lawlor. Sub-band independent subspace analysis for drum transcription. In *Proc. Digital Audio Effects Workshop (DAFx)*, 2002.

[Fra82] Paul Fraisse. Rhythm and tempo. In Deutsch [Deu82], pages 149–180.

[FU01] Jonathan Foote and Shingo Uchihashi. The beat spectrum: a new approach to rhythm analysis. In *Proc. Int. Conf. on Multimedia and Expo (ICME)*, 2001.

[Gab99] Alf Gabrielsson. The performance of music. In Deutsch [Deu99], pages 501–602.

[GFB03] Fabien Gouyon, Lars Fabig, and Jordi Bonada. Rhythmic expressiveness transformations of audio recordings: swing modifications. In *Proc. Digital Audio Effects Workshop (DAFx)*, 2003.

[GH03a] Fabien Gouyon and Perfecto Herrera. A beat induction method for musical audio signals. In *Proc. 4th WIAMIS Special Session on Audio Segmentation and Digital Music*, QM, Univ. London, 2003.

[GH03b] Fabien Gouyon and Perfecto Herrera. Determination of the meter of musical audio signals: seeking recurrences in beat segment descriptors. In *AES 114th Convention*, Amsterdam, March 2003.

[GHC02] Fabien Gouyon, Perfecto Herrera, and Pedro Cano. Pulse dependent analyses of percussive music. In *Proc. 22nd Int. Conf. on Virtual, Synthetic and Entertainment Audio*, 2002.

[GM97] Masataka Goto and Yoichi Muraoka. Issues in evaluating beat tracking systems. In *International Joint Conference on Artificial Intelligence, Workshop on Issues in AI and Music, Evaluation and Assessment*, pages 9–16, 1997.

[GM99] Masataka Goto and Yoichi Muraoka. Real-time beat tracking for drumless audio signals: chord change detection for musical decisions. *Speech Communication*, 27:311–35, 1999.

[GM03] Fabien Gouyon and Benoît Meudic. Towards rhythmic content processing of musical signals: Fostering complementary approaches. *Journal of New Music Research*, 32(1):41–64, 2003.

[Got01] Masataka Goto. An audio-based real-time beat tracking system for music with or without drum-sounds. *Journal of New Music Research*, 30(2):159–71, 2001.

[GT99] Niall Griffith and Peter M. Todd, editors. *Musical Networks: Parallel Distributed Perception and Performance.* MIT Press, Camb, MA, 1999.

[Gue03] Carlos Guedes. Controlling musical tempo from dance movement in real-time: A possible approach. In *Proc. Int. Computer Music Conference*, pages 453–7, 2003.

[Hai04] Stephen W. Hainsworth. *Techniques for the Automated Analysis of Musical Audio.* PhD thesis, University of Cambridge, 2004.

[Han89] Stephen Handel. *Listening: an introduction to the perception of auditory events*. MIT Press, Camb, MA, 1989.

[Har97] William M. Hartmann. The physical description of signals. In *Hearing* [Moo97a], pages 1–40.

[Has97] Christopher F. Hasty. *Meter* as *Rhythm*. Oxford University Press, New York, 1997.

[HM03] Stephen Hainsworth and Malcolm Macleod. Onset detection in musical audio signals. In *Proc. Int. Computer Music Conference*, pages 163–6, 2003.

[Hon01] Henkjan Honing. From time to time: The representation of timing and tempo. *Computer Music Journal*, 25(3), 2001.

[HWC91] Peter Howell, Robert West, and Ian Cross, editors. *Representing Musical Structure*. Academic Press, London, 1991.

[IBWW97] Vijay Iyer, Jeff Bilmes, Matt Wright, and David Wessel. A novel representation for rhythmic structure. In *Proc. Int. Computer Music Conference*, 1997.

[Imp01] Jonathan Impett. *Computational Models for Interactive Composition/Performance Systems*. PhD thesis, University of Cambridge, 2001.

[JA03] Kristoffer Jensen and Tue Haste Andersen. Real-time beat estimation using feature extraction. In *Proc. Computer Music Modeling and Retrieval Symposium, Lecture Notes in Computer Science*. Springer Verlag, 2003.

[JG03] Petr Janata and Scott T Grafton. Swinging in the brain: shared neural substrates for behaviors related to sequencing and music. *Nature Neuroscience*, 6(7), July 2003.

[JL91] Philip N. Johnson-Laird. Jazz improvisation: A theory at the computational level. In Howell et al. [HWC91], pages 291–325.

[JMF02] Tristan Jehan, Tod Machover, and Mike Fabio. Sparkler: An audio-driven interactive live computer performance for symphony orchestra. In *Proc. Int. Computer Music Conference*, Göteborg, Sweden, 2002.

[Jor02] Sergi Jorda. Improvising with computers: A personal survey (1989–2001). *Journal of New Music Research*, 31(1):1–10, 2002.

[JY93] Mari Riess Jones and William Yee. Attending to auditory events: the role of temporal organisation. In McAdams and Bigand [MB93], pages 69–112.

[Kla03] Anssi P. Klapuri. Musical meter estimation and music transcription. In *Proc. Cambridge Music Processing Colloquium*, pages 40–5, 2003.

[Lar00] Edward W Large. On synchronizing movements to music. *Human Movement Science*, 19:527–66, 2000.

[Lee91] Christopher S. Lee. The perception of metrical structure: Experimental evidence and a model. In Howell et al. [HWC91], pages 59–127.

[Lig98] György Ligeti. *Études pour piano: deuxième livre*. Schott, Mainz, Germany, 1998.

[LJ83] Fred Lerdahl and Ray Jackendoff. *A Generative Theory of Tonal Music*. MIT Press, Camb, MA, 1983.

[LJ99] Edward W. Large and Mari Riess Jones. The dynamics of attending: How people track time-varying events. *Psychological Review*, 106(1):119–59, 1999.

[LK94] Edward W. Large and John F. Kolen. Resonance and the perception of musical meter. *Connection Science*, 6:177–208, 1994.

[LM03] P. A. Lewis and R. C. Miall. Brain activation patterns during measurement of sub- and supra-second intervals. *Neuropsychologia*, 41:1583–92, 2003.

[LP02] Edward W. Large and Caroline Palmer. Perceiving temporal regularity in music. *Cognitive Science*, 26:1–37, 2002.

[LU97] Fa-Long Luo and Rolf Unbehauen. *Applied Neural Networks for Signal Processing.* Cambridge University Press, Cambridge, 1997.

[Mac03] David J. C. MacKay. *Information Theory, Inference, and Learning Algorithms.* Cambridge University Press, Cambridge, 2003.

[Mar81] Stephen M. Marcus. Acoustic determinants of perceptual center (p-center) location. *Perception and Psychophysics*, 30(3):247–56, 1981.

[Mar00] Alan Marsden. *Representing Musical Time: A Temporal-Logic Approach.* Swets and Zeitlinger, Lisse, the Netherlands, 2000.

[MB93] Stephen McAdams and Emmanuel Bigand, editors. *Thinking in Sound: The Cognitive Psychology of Human Audition.* Clarendon Press, Oxford, 1993.

[McA93] Stephen McAdams. Recognition of sound sources and events. In McAdams and Bigand [MB93], pages 146–98.

[McC02] James McCartney. Rethinking the computer music language: SuperCollider. *Computer Music Journal*, 26(4), 2002.

[Mic00] A Michon, John. Tapping and synchronization: Introduction. In Desain and Windsor [DW00], pages 83–94.

[MLM+03] Marco Molinari, Maria G. Legio, Martina De Martin, Antonio Cerasa, and Michael Thaut. Neurobiology of rhythmic motor entrainment. In *Annals New York Academy of Sciences 999*, pages 313–21, 2003.

[Moo97a] Brian C. J. Moore. *Hearing.* Academic Press, San Diego, CA, 1997.

[Moo97b] Brian C. J. Moore. *An Introduction to the Psychology of Hearing.* Academic Press, San Diego, CA, 1997.

[MP97] Jonathan M. Magill and Jeffrey L. Pressing. Asymmetric cognitive clock structures in West African rhythms. *Music Perception*, 15(2):189–222, 1997.

[MRMP94] Jiří Mates, Tomáš Radil, Ulrike Müller, and Ernst Pöppel. Temporal integration in sensorimotor synchronization. *Journal of Cognitive Neuroscience*, 6(4):332–40, 1994.

[MSV98] K.D. Martin, E.D. Scheirer, and B.L. Vercoe. Music content analysis through models of audition. In *ACM Workshop on Content Processing of Media for Multimedia Applications, Bristol*, 1998.

[MSWW01] Timothy Madden, Ronald B. Smith, Matthew Wright, and David Wessel. Preparation for interactive live computer performance in collaboration with a symphony orchestra. In *Proc. Int. Computer Music Conference*, Havana, Cuba, 2001.

[OD01] Nicola Orio and François Déchelle. Score following using spectral analysis and hidden Markov models. In *Proc. Int. Computer Music Conference*, Havana, Cuba, 2001.

[Pal97] Caroline Palmer. Music performance. *Annual Review of Psychology*, 48:728–41, 1997.

[Par94] Richard Parncutt. A perceptual model of pulse salience and metrical accent in musical rhythms. *Music Perception*, 11(4):409–64, 1994.

[Par99] Joseph Paradiso. The brain opera technology: New instruments and gestural sensors for musical interaction and performance. *Journal of New Music Research*, 28(2):130–49, 1999.

[PAZ98] Miller S. Puckette, Theodore Apel, and David D. Zicarelli. Real-time audio analysis tools for Pd and MSP. In *Proc. Int. Computer Music Conference*, pages 109–112, 1998.

[PB02] Bryan Pardo and William Birmingham. Improved score following for acoustic performances. In *Proc. Int. Computer Music Conference*, Göteborg, Sweden, 2002.

[PE85] Dirk-Jan Povel and Peter Essens. Perception of temporal patterns. *Music Perception*, 2(4):411–440, 1985.

[PHH99] Joseph Paradiso, Kai-Yuh Hsiao, and Eric Hu. Interactive music for instrumented dancing shoes. In *Proc. Int. Computer Music Conference*, Beijing, China, 1999.

[PK03] Jouni Paulus and Anssi Klapuri. Model-based event labelling in the transcription of percussive audio signals. In *Proc. Digital Audio Effects Workshop (DAFx)*, 2003.

[PRK01] Arkady Pikovsky, Michael Rosenblum, and Jürgen Kurths. *Synchronization: A Universal Concept in Nonlinear Sciences*. Cambridge University Press, Cambridge, 2001.

[PW99] Ernst Pöppel and Marc Wittman. Time in the mind. In Robert A. Wilson and Frank Keil, editors, *The MIT Encyclopedia of the Cognitive Sciences*, pages 841–3. MIT Press, Camb, MA, 1999.

[Rap01a] Chris Raphael. Automated rhythm transcription. In *Proc. Int. Symp. on Music Information Retrieval*, 2001.

[Rap01b] Chris Raphael. Synthesizing musical accompaniments with Bayesian belief networks. *Journal of New Music Research*, 30(1):59–67, 2001.

[Rap04] Chris Raphael. Aligning musical scores with audio using hybrid graphical models. *American Association for Artificial Intelligence*, 2004. Forthcoming, http://fafner.math.umass.edu/papers/index.html.

[Rep00a] Bruno Repp. Rhythm and meter in music and speech: Introduction. In Desain and Windsor [DW00], pages 235–7.

[Rep00b] Bruno Repp. Subliminal temporal discrimination revealed in sensorimotor coordination. In Desain and Windsor [DW00], pages 129–142.

[Rep01] Bruno Repp. Processes underlying adaptation to tempo changes in sensorimotor synchronisation. *Human Movement Science*, 20:277–312, 2001.

[Roa96] Curtis Roads. *The Computer Music Tutorial*. MIT Press, Cambs, MA, 1996.

[Ros92] David Rosenthal. Emulation of human rhythm perception. *Computer Music Journal*, 16(1):64–76, 1992.

[Row93] Robert Rowe. *Interactive Music Systems*. MIT Press, Cambs, MA, 1993.

[Row01] Robert Rowe. *Machine Musicianship*. MIT Press, Cambs, MA, 2001.

[RRG99] Geber L. Ramalho, Pierre-Yves Rolland, and Jean-Gabriel Ganascia. An artificiality intelligent jazz performer. *Journal of New Music Research*, 28(2):105–29, 1999.

[Rya91] Joel Ryan. Some remarks on musical instrument design at STEIM. *Contemporary Music Review*, 6(1):3–17, 1991.

[Sch98] Eric D. Scheirer. Tempo and beat analysis of acoustic musical signals. *J. Acoust. Soc. Am.*, 103(1):588–601, January 1998.

[Sch99] Eric D. Scheirer. Towards music understanding without separation: Segmenting music with correlogram comodulation. In *IEEE Workshop on Applications of Signal Processing to Audio and Acoustics*, 1999.

[Sep01] Jarno Seppänen. Computational models of musical meter recognition. Master's thesis, Dept. Information Technology, Tampere University of Technology, August 2001.

[SFF91] Johan Sundberg, Anders Friberg, and Lars Frydén. Common secrets of musicians and listeners: An analysis-by-synthesis study of musical performance. In Howell et al. [HWC91], pages 161–97.

[SFvC00] Ricarda I. Schubotz, Angela D. Friederici, and D. Yves von Cramon. Time perception and motor timing: a common cortical and subcortical basis revealed by fMRI. *NeuroImage*, 11:1–12, 2000.

[Slo82] John A. Sloboda. Music performance. In Deutsch [Deu82], pages 479–496.

[Smi96] Leigh M. Smith. Modelling rhythm perception by continuous time-frequency analysis. In *Proc. Int. Computer Music Conference*, Hong Kong, 1996.

[SS01] William A. Sethares and Thomas W. Staley. Meter and periodicity in musical performance. *Journal of New Music Research*, 30(2):149–58, 2001.

[TB96] Neil P. McAngus Todd and Guy J. Brown. Visualisation of rhythm, time and metre. *Artificial Intelligence Review*, 10:253–73, 1996.

[TEC02] George Tzanetakis, Georg Essl, and Perry Cook. Human perception and computer extraction of beat strength. In *Proc. Digital Audio Effects Workshop (DAFx)*, 2002.

[Tem01] David Temperley. *The Cognition of Basic Musical Structures.* MIT Press, Camb, MA, 2001.

[Tho00] Belinda Thom. Artificial intelligence and real-time interactive improvisation. In *AAAI Music and AI Workshop*, Austin,TX, 2000.

[TL91] Peter M. Todd and D. Gareth Loy, editors. *Music and Connectionism.* MIT Press, Camb, MA, 1991.

[TLO02] Neil P. McAngus Todd, Christopher S. Lee, and Donald J. O'Boyle. A sensorimotor theory of temporal tracking and beat induction. *Psychological Review*, 66:26–39, 2002.

[Tod92] Neil P. McAngus Todd. The dynamics of dynamics: a model of musical expression. *J. Acoust. Soc. Am.*, 91(6):3540–50, 1992.

[Toi98] Petri Toiviainen. An interactive MIDI accompanist. *Computer Music Journal*, 22(4):63–75, Winter 1998.

[Toi01] Petri Toiviainen. Real-time recognition of improvisations with adaptive oscillators and a recursive Bayesian classifier. *Journal of New Music Research*, 30(2):136–46, 2001.

[TOL99] Neil P. McAngus Todd, Donald J. O'Boyle, and Christopher S. Lee. A sensory-motor theory of rhythm, time perception and beat induction. *Journal of New Music Research*, 28(1):5–28, 1999.

[TS00] Petri Toiviainen and Joel Synder. The time course of pulse sensation: Dynamics of beat induction. In *Proc. Int. Conf. on Music Perception and Cognition*, Keele, 2000.

[UH03] Christian Uhle and Juergen Herre. Estimation of tempo, micro time and time signature from percussive music. In *Proc. Digital Audio Effects Workshop (DAFx)*, 2003.

[Ver97] Barry Vercoe. Computational auditory pathways to music understanding. In Irène Deliège and John Sloboda, editors, *Perception and Cognition of Music*, pages 307–26. Psychology Press, Hove, UK, 1997.

[Ver00] Barry Vercoe. Understanding Csound's spectral data types. In Richard Boulanger, editor, *The Csound Book*, pages 437–47. MIT Press, Camb, MA, 2000.

[vNM99] Leon van Noorden and Dirk Moelants. Resonance in the perception of musical pulse. *Journal of New Music Research*, 28(1):43–66, 1999.

[VP93] Neal F. Viemeister and Christopher J. Plack. Time analysis. In William A. Yost, Arthur N. Popper, and Richard R. Fay, editors, *Human Psychophysics*, pages 116–54. Springer, New York, 1993.

[VR81] Joos Vos and Rudolf Rasch. The perceptual onset of musical tones. *Perception and Psychophysics*, 29(4):323–35, 1981.

[VW96] Dirk Vorberg and Alan Wing. Modeling variability and dependence in timing. In Herbert Heuer and Steven W. Keele, editors, *Handbook of Perception and Action, Volume Two: Motor Skills*, pages 181–262. Academic Press, London, 1996.

[Wal97] William F. Walker. A computer participant in musical improvisation. In *Proc. Computer-Human Interaction (CHI)*, 1997.

[WC04] Ge Wang and Perry R. Cook. On-the-fly programming: Using code as an expressive musical instrument. In *New Interfaces for Musical Expression (NIME)*, Hamamatsu, Japan, 2004.

[Wei93] Arthur Weisberg. *Performing Twentieth Century Music: a Handbook for Conductors and Instrumentalists.* Yale University Press, New Haven, 1993.

[WF97] Matthew Wright and Adrian Freed. Open Sound Control: A new protocol for communicating with sound synthesisers. In *Proc. Int. Computer Music Conference*, pages 101–4, Thessaloniki, Hellas, 1997.