Towards a Style-Specific Basis for Computational Beat Tracking

Nick Collins
Centre for Music and Science,
Faculty of Music, Cambridge
nc272@cam.ac.uk

ABSTRACT

Outlined in this paper are a number of sources of evidence, from psychological, ethnomusicological and engineering grounds, to suggest that current approaches to computational beat tracking are incomplete. It is contended that the degree to which cultural knowledge, that is, the specifics of style and associated learnt representational schema, underlie the human faculty of beat tracking has been severely underestimated. Difficulties in building general beat tracking solutions, which can provide both period and phase locking across a large corpus of styles, are highlighted. It is probable that no universal beat tracking model exists which does not utilise a switching model to recognise style and context prior to application.

Keywords

Beat tracking, metre perception, re-synchronisation

Note

The author acknowledges that much of this work is speculative, and tries to suggest areas where future experiments may help to resolve matters. There are intimate ties to certain issues in signal processing, particularly transcription and auditory scene analysis, that there is not room to fully explore. Beat tracking in this article means the determination of both period and phase (tempo and exact beat placement). For reasons of space I have also left out certain signal processing and experimental details that will be more properly covered in my forthcoming PhD thesis, to be submitted summer 2006.

Proceedings of the 9th International Conference on Music Perception & Cognition (ICMPC9).

EVIDENCE FROM PSYCHOLOGY AND ETHNOMUSICOLOGY

`Anticipation can only come from familiarity with the norms of a musical style, providing another motivation for beginning to include a base of musical knowledge in computer programs.' (Rowe 1993, p117)

Musicians are often surprised when the difficulty of programming a computer to tap along to the beat of music is mentioned to them. To any adult participant in a given culture, moving in time with their music is so natural an ability that it is easy to forget that it is not a given in early development, but usually becomes established by the age of four (Drake et al. 2000, Drake and Bertrand 2003, McAuley and Jones 2005).

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' (Clayton 2000, p 41). He notes that the well-formedness rules for metre of Lerdahl and Jackendoff's theory (1983) cannot accommodate North Indian tal 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 Stobart and Cross (2000). They study downbeat location in the music of the Northern Potosí of Bolivia, finding that what seems to an anacrusis to the authors' Western training is in fact the downbeat itself. Temperley (2001), 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 can be fruitful.

Yet there are at least three theories of metre concerning African rhythm. Arom (1989) finds an isoperiodic pulse level and subdividing operational value (tatum) 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 (1995) meanwhile argues for a conventional metrical backdrop to the Northern Ewe music of Ghana. Magill and Pressing (1997) describe the nonisochronous *timeline* as the the best fit for a West African drummer's mental model of polyrhythmic production.

A resolution of such a proliferation of metrical theories is explained by Justin London's *many-metres hypothesis* (London 2004) which contends that there are a multiplicity of both isochronous and non-isochronous (though pseudo-isochronous) organised metrical structures, each of which must be learnt in a given context. The same I-metre or NI-metre may also vary in its cognitive evocation as tempo is varied, due to the interaction of processing considerations for the different levels of pulsation, particularly with regard to a fundamental minimum treatable IOI unit around 100 milliseconds, and the need for beats themselves to be at least 250 msec.

Some non-isochronous metres may have arisen through the placing of uneven dance steps, a short-short-long (perhaps a step-step-turn) corresponding to a 2+2+3, or contentiously dropping any insistence on quantisation, 2+2+2.9. An underlying tatum (subdivisor, fastest operational unit) is not necessarily implied in a short-long dance steps view, which would undermine the cognitive existence of London's n-cycle, though not of a master periodicity per se. Norwegian Hardanger fiddle music as well as much Aksak (Eastern European) dance music show these `choreographic' (Brailioiu 1984, p136) tendencies. The weakest part of London's theory is possibly related to the NI-meters, which are usually maximally evenly spaced within an n-cycle even if they cannot be made perfectly isochronous. Examples like Hardanger fiddle show that listeners may simply be familiar with certain patterns with larger scale periodicity, subdivided in a way intimately tied to bodily motion in the dance, though not necessarily mathematically neat. The influence of Western music theoretic constructions does tend to apply quantisation, or force the positing of an n-cycle or tatum, where this may not be necessary to explain the phenomena, and some of the controversy is shown by the competing characterisations of certain African rhythmic constructs mentioned above.

Thaut (2005, p54) accepts the validity of multiple metrical constructs in explaining the world's rhythms: `Because very different neurological networks, which seem to be fairly independent of each other in circuit architecture, subserve different components of rhythm, hierarchical rhythmic structures, such as those in Western musical language systems, may be based more on the particular structural developments in the syntax of musical languages specific to certain musical cultures than on a culturally independent intrinsic function of a musical biology'. It may be possible to claim that there is no underlying isochronous pulse or tatum in some music, just as some Indian tal admit a pulse, but involve much longer cycles than typical Western metres. Whilst Drake and Bertrand (2003) posit musical temporal universals, in particular the *predisposition towards regularity* and *active search for regularity*, a more general version of such rules may have to be allowed which does not assume isochrony at a short time scale, but larger scale regularity of reference pattern. Further cross-cultural study is needed to resolve the status of pulse especially, and brings us to a consideration of scheduling and perceptual clock mechanisms. If the outer periodicity can be long, how many reference points might be needed in a cycle for musical timing purposes? Must these be isochronous, or be constrained to simple integer ratios?

Cultural conventions and stylistic differences in the treatment of metre and pulse seem to be corroborated by the psychological literature and ethnomusicological accounts. In further reports, Snyder et al. (2005) show that familiarity is important for metre tracking in a tapping study on Balkan non-isochronous meter. Jehan (2005), in a computational study, recently demonstrated the utility of machine learning based on event timbres for downbeat induction in a context, Brazilian *Maracatu* dances, where many Western listeners were unable to find the appropriate ground truth. Such culturally specific conventions are a clue that we have no innate general beat tracking facility, and that training is required to resolve particular metrical structures. Since the resolution of a beat level may co-depend on aspects of the metrical formation and basic patterns, in turn influenced by instrumentation and other timbral markers, phase determination for computational beat trackers is hostage to potentially high level cultural factors.

EVIDENCE FROM OBSERVATION MODELS

Scheirer (1998), in a much cited paper, makes the claim that amplitude modulated noise in six bands can evoke the same beat percept as the original signal. This claim is taken as evidence to support the general utility and psychoacoustic relevance of his computational beat tracker, based on a similar signal processing frontend. Like most current generation trackers, the observation model is attuned to changes in energy features, in a limited number of bands. As reported at RPPW last year (Collins and Cross 2005) an experiment was run to reproduce Scheirer's study, by contrasting subject performance on both original CD quality and vocoded versions of real polyphonic music signals in a tapping paradigm. Data was sought on synchronisation accuracy, with a subsidiary interest in registered reaction times.

This experiment demonstrated a statistically significant change in performance of beat tracking across signal qualities, with Scheirer's six-band vocoding reducing the abilities of subjects to synchronise effectively. It is thus contended that Scheirer's multi-band frontend approach on its own is not sufficient to model human musical ability. More advanced and integrated timbral information of individual events is a contender for the true mechanism by which humans learn and respond to music; six bands is certainly a simplification of the rich information channels from 3500 inner hair cells!

It has actually been recognised by engineers that such short-term energy change frontends are most appropriate to percussive music, and indeed, alternative mechanisms have been suggested to cope with situations where longer range tonal information may be more important than transient cues. Goto (2001) tried to take chordal information into account for `music without drum-sounds' when rating beat tracking hypotheses, and Hainsworth (2004) uses a longer window chord detection principle as a complement to percussive event detectors, in particular for the consideration of choral and classical music; both of these authors, however, run the chord detection in parallel with percussive onset detection processes. Davies and Plumbley (2005b) compared frontends to a common autocorrelation model. The most successful was founded in the complex domain onset detection principle, which utilises tonal information implicitly via instantaneous frequency (change of phase in FFT bins) concurrently with percussive (change of amplitude). Interestingly, by allowing the detection function to be genre, piece or even frame specific, Davies could improve the tracker performance on a 100 piece, 5 genre corpus (a subset of the Hainsworth database) by 15\%. The author, who converted the Davies model to a real-time implementation for concert use (a SuperCollider UGen called AutoTrack) also found that the complex domain frontend was superior to a pure energy transient based model. However, neither the non-causal non-realtime Davies model nor AutoTrack could be made to perform better than an overall F-measure of 66% and 47% respectively over the 200 piece Hainsworth beat tracking test database, on stringent conditions of beat matching. Whilst drum-heavy genres were effectively tracked, performance tailed off for classical and folk idioms, suggesting that further work would have to be done to tackle these contexts. A general observation model may not support all styles.

The issue of the best frontend has been raised by Gouyon (2005) with an exhaustive comparison of framewise features and possible algorithms for the integration of information arising from feature sets, by combinations both before and after periodicity detection functions are generated. Unfortunately, Gouyon's evaluation is cast in terms of tempo induction, and does not consider beat. This is problematic, for the determination of the phase is perhaps the most critical facility of human beat tracking required for musical interaction. He also considers features that may best relate to beats as predictors, considering feature values on beats and non-beats over a large annotated database. The single best feature varies with genre, corroborating Davies' work and common sense, Gouyon (2005, p99) admitting `different sets of features would probably be relevant depending on the musical style'. It might be argued that the use of single frame features (and their first order differences implicitly involving two successive frames) is not so perceptually relevant as the scale of eventwise features, timbral sound objects and their role in a stylistic context, which are only indirectly acknowledged via the frame values.

Tristan Jehan's aforementioned method of downbeat induction (Jehan 2005) uses such segments discovered via a Bark subband frontend. He provides an early example of a machine learning study on resolving the location of the downbeat, and his method is inherently style-specific. It could be possible that downbeat estimation is the special stage requiring training, and that discovery of a basic beat level can be carried out by standard general signal energy methods. Yet the psychological aspects of metre would suggest situations where beat and measure, or master periodicity and timeline pattern, must be co-estimated from the observations.

The author's experience, in preparing systems for concert use, has engendered a strong bias to particularisation, in specialising systems to the task they face. General beat tracking models may be usable, but can be improved by giving them frontends appropriate to the situation. In evaluating a recent system designed to track a baroque duo of recorder and harpsichord, and which had specialised onset detection mechanisms for those instruments, the Davies non-realtime beat tracker was improved by 16% and the AutoTrack realtime beat tracker by 4% in evaluations by using the specialised observation model, rather than the complex domain onset detection.

EVIDENCE FROM RE-SYCHRONISATION TIME

Periodicity detection mechanisms in beat trackers tend to utilise some form of search procedure, commonly via an autocorrelation or comb filtering (Gouyon 2005, p104), on a 3-6 second window of feature data. Ideally assuming no tempo deviation within this window, an induction of the period is more stable for longer windows: the Davies and Plumbley (2005a) model uses a six second window, covering 8 beats at 4/4 for many tempi. The disadvantage of larger windows is a slower reaction time, and this may itself provide evidence that the assumed mechanisms of beat tracking are distinctly non-human. In experiments on re-sychronisation time reported last year (Collins and Cross 2005) reaction times of 1-2 seconds were commonly seen. Further, by rating cases where subjects could be said to provide a valid re-sychronisation at all, humans significantly outperformed existing beat tracking models on an `obstacle course' of piece transitions. By splicing together varied piece extracts in 6 second segments with jumps of phase that forced realignment and differing tempi to confound tempo priors, an evaluation method for computational beat trackers was exhibited which clearly differentiated computer from human.

Table 1

Subject	**Mean Score**	**Mean Reaction Time**	**Successes**
Beat Musician	0.308	1.369	19
Average Subject	0.604	1.686	19
Davies and Plumbley (2005a) non-causal	0.737	0.956	14
Klapuri et al. (2006) non-causal	0.756	1.788	13
Klapuri et al. (2006) causal	0.815	2.094	13
Scheirer (1998)	0.886	2.278	10
AutoTrack	1.168	2.697	12

Table 1 gives a comparison of human and beat tracker performance on the experimental test set. Computational beat trackers well-known in the literature ran the obstacle course. There were 20 extracts spliced together, with 19 transitions, and measurements were given after each transition of how well a subject resynchronised (the score, lower scores being best) and how quickly (the reaction time). The last column is critical, in that it shows how many of the transitions were accurately responded to.

The computational algorithms are slower to resynchronise, and fail to re-sychronise after some transitions. A few reaction times are rather suspicious- it is quite possible that a few transitions may have been coped with by a lucky prior state of the tracker. Non-causal algorithm reaction times are of course to be taken with a pinch of salt.

It would be unfair to say that in reaction time humans are greatly faster than computational beat trackers, but humans are certainly more reliable, even non-musicians far outperfoming the computer models.

At the very least, perhaps humans have an advantage in spotting the transitions themselves, perhaps through recognising characteristic instruments, playing styles and misalignments of key or other timbral-harmonic features. This is a reminder of the `radio-dial' paradigm explored by Perrot and Gjerdingen (1999) where subjects could distinguish genre even from half second extracts. Hainsworth (2004) has suggested timbre perception as an essential early component of any transcription model, and Koelsch (2005), in a neuroscientific review, places timbre determination early on in auditory processing (within 100 milliseconds).

How could this be effected computationally? Following the relatively crude framewise (10 millisecond spaced FFT frames) features typical of music information retrieval work, broad timbral measures accumulated from those frames might distinguish sections. An overall measure of harmonic information and timbre by spotting novel spectral content in non-transient regions was utilised to create a detection function with a three second delay in causal operation. Re-initialisation of a beat tracker was forced by positive detections peak picked from this function. The full signal processing details are in my forthcoming PhD thesis, to be submitted this summer.

Table 2

Subject	**Mean Score**	**Mean Reaction Time**	**Successes**
Davies and Plumbley (2005a) non-causal given all transitions	0.585	0.2	16
Davies and Plumbley (2005a) non-causal given discovered transitions	0.666	0.52	15

The Davies model was adapted to reset the current working period hypothesis and phase prior at selected beat induction steps, namely, at those points given by transition data. Table 2 shows a small improvement in tracking performance for the Davies beat tracking algorithm when it is apprised of transition locations in taking the obstacle course test. Even with perfect knowledge of transitions, the algorithm still fails to show adequate reaction to three transitions. This is no doubt because the significant events within these sections are not revealed by the onset detection frontend, and forcing a beat tracker recalculation at the appropriate moment will not change the frontend's signal processing capabilities.

It could be argued that some difficulties of detecting appropriate events in the beat tracker frontend are also difficulties of the event analysis implicit in the transition detector. The relative simplicity of the transition detection process can be linked to the relative simplicity of the audio signal analysis on which the beat tracker operates, compared to the clarity of a human's understanding of objects relevant to beat perception. Improvements in observation frontends, prioritised rather than tracking models, seem to be demanded by the obstacle course test.

Whilst a great improvement to the beat tracking commensurate with human performance has not been exhibited, the fact that tracker performance could be improved at all is sufficient justification for the investigation. It is highly apposite if it is accepted that style-specific knowledge is essential for beat tracking to consider mechanisms for identifying stylistic character early on in processing, so as to select a specialised onset detection frontend and beat tracker appropriate to the task. It would be pertinent to consider what the music information retrieval community could provide to this sort of multistage algorithm.

There are also similarities between this work and the goals of MIR in the analysis of sections within a composition, and music similarity measures between pieces based in various audio descriptors (Pampalk 2004, Berenzweig at al. 2003, Aucouturier 2002). However, the application described here is more like a between-piece differentiator suitable for online application. The transition detector must trigger if the radio dial has just been jogged; the obstacle course is not a single realistic piece of music in itself, and the transitions are not known a priori. It may be possible to adapt more advanced piece discriminating functions from the MIR literature to compare small windows of audio either side of a potential transition, in a causal fashion, though some music similiarity measures dependent on statistics across whole pieces will not be appropriate for spotting transitions between short extracts.

The extent to which framewise statistical methods could fully solve an `obstacle course' problem remains to be seen. The author's intuition is that eventwise (composite framewise) rather than simple framewise features may prove most apposite to matching human performance, our learnt sound objects and their horizontal and vertical interactions in patterns being implicated in human beat tracking. There is no conclusive proof that a low-level feature basis with machine learning could not cope, however, and events are implicit in the combination of feature frames; perhaps the most important contention is that longer term properties (continuities) of sounds than 20 milliseconds are to be taken into account (Kapanci and Pfeffer 2004, Collins 2005). However, by even implicitly raising the relation of style to beat tracking performance, we are forced to accept a greater role for learnt schema in the process.

CONCLUSIONS

Beat tracking models with adaptive frontends may form the next generation of computational systems. Such a dependence on context is indicative of critical timbral-stylistic factors in beat tracking, as opposed to an unrealistic expectancy of a general beat tracking solution running from a single universal frontend.

Speculatively, perhaps the human superiority over state-of-the-art algorithms is due to an eventwise rather than instantaneous framewise formulation of signal features? If the recognition of context is essential, the separation of streams based on instrument, and the re-integration of instrumental lines based on style may provide a better model. Styles may be indicated by an aggregate of timbral cues relatively early on in processing, assisting the selection of prior, and explaining the fast reactions of human subjects for re-sychronisation after discontinuity or as musical performers in ensemble interactions. Knowing that some events are designations of the downbeat, that the enclosing periodicity of a cycle is marked out by particular patterns on particular instruments or combinations of instruments is crucial to fast responses, for as long as enough context is granted, one can infer the existence of the rest. A partial match of markers to an instrumental situation is sufficient to then predict beat locations. An eventwise view, requiring stream separation and instrument recognition (even without assuming perfect extraction), makes demands on the signal processing technology that go far beyond existing onset detection frontends employed in current beat trackers. And where onset detection itself can be shown to utilise learnt categorisations (Windsor 2000, Bregman 1990), the style dependence is only increased.

How might we further investigate such contentions experimentally? Aside from following up (Collins and Cross 2005) with larger scale experiments, one might imagine the comparison of pieces built without timbral cues with the full event information. These could be synthesised via scores either with a set of appropriate instruments, or just as an aggregate pattern on a single instrument. One could then measure subject tapping behaviour, in particular, the ambiguity of the location of the downbeat and the appropriate (stylistic conventional) metrical level. As reproducing the Scheirer experiment suggested, patterns without sufficiently discriminating characteristics for events become more difficult to resolve. In particular, such experiments can highlight the inadequacy of generalising from tapping studies with bare (and usually metronomic, isochronous) stimuli to polyphonic audio.

To attempt to answer London's (2004, p158) question of how we learn meters: perhaps we tag meter types with the instruments typically involved. This might extend the `many meters hypothesis' to further include timbral factors critical to stylistic familiarity as differentiating metrical settings. There may be experimental tests using non-standard arrangements.

In this paper I have tried to cast some doubt over the long window periodicity detection, and framewise feature frontends of current generation computational beat tracking models, from a number of sources of evidence. If we are to accept the multiplicity of metrical constructs in the world, we must model the training encultured listeners undergo in recognising and synchronising with contexts. In particular, the building of machine musicians able to interact in real-time performance is assisted by a practical specialisation in styles, without the expectation of universal beat tracking solutions.

`The human construct that we call our *music* is merely a convention- something we have all evolved together, and that rests on no final or ultimate laws.' (Reich 2002, p131)

ACKNOWLEDGMENTS

I would like to thank my colleagues in the CMS for fruitful discussions, and the various beat tracking researchers I have dealt with, in particular Tristan Jehan, Matthew Davies and Anssi Klapuri, all of whom made code available. This work was supported by AHRC grant 2003/104481.

REFERENCES

Agawu, V. K. (1995). *African Rhythm: A Northern Ewe Perspective*. Cambridge: Cambridge University Press.

Arom, S. (1989). Time structure in the music of Central Africa: Periodicity, meter, rhythm and polyrhythmics.

Leonardo, 22(1), 91–9.

Aucouturier, J-J. (2002). Music similarity measures: What's the use? In *Proc. Int. Symp. on Music Information Retrieval*.

Berenzweig, A., Logan, B., Ellis, D., & Whitman, B. (2003). A large-scale evaluation of acoustic and subjective music similarity measures. In *Proc. Int. Symp. on Music Information Retrieval*.

Brailoiu, C. (1984). *Problems of Ethnomusicology*. Cambridge: Cambridge University Press.

Bregman, A. S. (1990). *Auditory Scene Analysis: The Perceptual Organization of Sound*. Camb, MA: MIT Press.

Clayton, M. (2000). *Time in Indian Music: Rhythm, Metre and Form in North Indian Rag Performance*. Oxford: Oxford University Press.

Collins, N. (2005). Using a pitch detector as an onset detector. In *Proc. Int. Symp. on Music Information Retrieval*.

Collins, N., & Cross, I. (2005). Beat tracking and reaction time. In *Rhythm Perception and Production Workshop (RPPW10)*.

Davies, M. E. P., & Plumbley, M. D. (2005a). Beat tracking with a two state model. In *Proceedings of IEEE Int. Conf. on Acoustics, Speech, and Signal Processing*.

Davies, M. E. P., & Plumbley, M. D. (2005b). Comparing mid-level representations for audio based beat tracking. In *Proceedings of the DMRN Summer Conference.*

Drake, C., & Bertrand, D. (2003). The quest for universals in temporal processing in music. In Peretz, I., & Zatorre, R. (Eds.), *The Cognitive Neuroscience of Music* (pp. 21–31). New York: Oxford University Press.

Drake, C., Jones, M. R., & Baruch, C. (2000). The development of rhythmic attending in auditory sequences: attunement, referent period, focal attending. *Cognition, 77*, 251–88.

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

Gouyon, F. (2005). *A Computational Approach to Rhythm Description: Audio Features for the Computation of Rhythm Periodicity Features and their use in Tempo Induction and Music Content Processing*. PhD thesis, Universitat Pompeu Fabra.

Hainsworth, S. W. (2004). *Techniques for the Automated Analysis of Musical Audio*. PhD thesis, University of Cambridge.

Jehan, T. (2005). Downbeat prediction by listening and learning. In *IEEE Workshop on Applications of Signal Processing to Audio and Acoustics*.

Kapanci, E., & Pfeffer, A. (2004). A hierarchical approach to onset detection. In *Proc. Int. Computer Music Conference.*

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

Koelsch, S. & Siebel, W. A. (2005). Towards a neural basis of music perception. *Trends in Cognitive Sciences, 9(12).*

Lerdahl, F. & Jackendoff, R. (1983). *A Generative Theory of Tonal Music*. Camb, MA: MIT Press.

London, J. (2004). *Hearing in Time: Psychological Aspects of Musical Meter*. New York: Oxford University Press.

Magill, J. M., & Pressing, J. L. (1997). Asymmetric cognitive clock structures in West African rhythms. *Music Perception, 15(2)*, 189–222.

McAuley, D., & Jones, M. R. (2005). Time of our lives I: Changes in preferred tempi and regions of entrainment across the lifespan. In *Entrainment Network Conference*.

Pampalk, E. (2004). A Matlab toolbox to compute music similarity from audio. In *Proc. Int. Symp. on Music Information Retrieval*.

Perrot, D., & Gjerdingen, R. O. (1999). Scanning the dial: An exploration of factors in the identification of musical style (abstract only). In *Proc. Int. Conf. on Music Perception and Cognition*.

Reich, S. (2002). Hillier, P. (Ed.) *Writings on Music 1965-2000*. New York: Oxford University Press.

Rowe, R. (1993). *Interactive Music Systems*. Cambs, MA: MIT Press.

Scheirer, E. D. (1998). Tempo and beat analysis of acoustic musical signals. *J. Acoust. Soc. Am., 103(1)*, 588–601.

Snyder, J. S., Hannon, E. E., Large, E. W., & Christiansen, M. H. Synchronization and continuation tapping to complex meters. In *Rhythm Perception and Production Workshop (RPPW10).*

Stobart, H. & Cross, I. (2000). The Andean anacrusis? Rhythmic structure and perception in Easter songs of Northern Potosi, Bolivia. *British Journal of Ethnomusicology, 9(2)*, 63–94.

Temperley, D. (2001). *The Cognition of Basic Musical Structures*. Camb, MA: MIT Press.

Thaut, M. (2005). *Rhythm, Music, and the Brain*. New York: Routledge.

Windsor, L. (2000). Through and around the acousmatic: The interpretation of electroacoustic sounds. In Emmerson, S. (Ed.), *Music, Electronic Media and Culture* (pp 7–35). Aldershot, England :Ashgate Publishing Limited.