# MACHINE LISTENING TO PERCUSSION: CURRENT APPROACHES AND FUTURE DIRECTIONS.

**Michael Ward**
Academy of Music, Sound & Performance.
Leeds Metropolitan University
m.j.ward@leedsmet.ac.uk

**ABSTRACT**

Many approaches have been taken to detect and classify percussive events within music signals for a variety of purposes with differing and converging aims. In this paper an overview of those technologies is presented and a discussion of the issues still to overcome and future possibilities in the field are presented. Finally a system capable of monitoring a student drummer is envisaged which draws together current approaches and future work in the field.

## 1. INTRODUCTION

Extraction of percussive instruments from music signals differs to that of pitched instruments due to their noisy or stochastic nature due to the multiple modes of excitement they exhibit [1] and techniques have been developed to extract information about the temporal locations and origin of sounds, indeed a model has been developed which conducts an analysis of the stochastic nature of the signal to determine percussive onsets [2]. Beyond onset detection (a review of which falls beyond the scope of this paper but can be found in [3] and [4]) of a percussive event, many models have been presented which seek to identify the instrument within a drum kit responsible for the onset known as classification. Many of the techniques discussed here fall into this process flow of onset detection, feature extraction and classification. Extraction of percussive information increases in difficulty as the signal becomes more musically complex and three broad challenges exist. Firstly, identification of isolated drum sounds is the simplest of these, and has been achieved with good levels of success typically using correlation approaches. Overlapping drum sounds and thirdly overlapping drum sounds in the presence of other instruments present further difficulties requiring more sophisticated techniques to achieve the aim. Another broad classification of models could be those operating in real-time versus non real time algorithms, and both with be discussed here. Firstly the current approaches and their research area are discussed. The future possibilities in these fields are then discussed in relation to those areas of research which remain unresolved.

## 2. CURRENT APPROACHES

### 2.1. Beat Tracking

Beat Tracking is the real-time extraction of the underlying pulse within a piece of music. This is very simple for human listeners but evidence such as subjective rhythm, expectation and beat continuing during silence shows that the creation of a machine capable of such a task is extremely difficult. While beats exist within a piece of music that does not contain percussion, several studies have used percussive models to inform a beat tracker where percussion exists, for example [5] and [6]. Such models focus on locating bass and snare drum onsets as these often have onset coincidence with metric levels of the beat. While not strictly a beat tracker, the Pd object bonk~ [7] is worth mentioning alongside these models due to its real time Bounded Q analysis to locate noisy events. This object can also match events with pre defined templates in an attempt to classify the instrument.

### 2.2. Score Following and Score Alignment

Score following is the ability of a machine to follow a performer with reference to a notated score. Score alignment is the non real-time equivalent. Focus in this field has been on pitched rather than percussive instruments although work in score alignment for percussion and improving onset location has been done for example by [8]. This model detects snare and bass drum events only using a combination of correlation and dynamic time warping to match templates to events. Real-time algorithms require compromise on accuracy versus computational expense while non real-time models can analyse the whole performance for greater accuracy, the application of which will be discussed later.

### 2.3. Automated Transcription

Automated Transcription seeks to resolve a musical signal into a symbolic notation. Human transcription is a learned, non realtime process involving multiple passes. Real-time models offering better that human performance are therefore extremely unlikely and again modelling human capabilities is a non trivial task. Expressive deviations which while musically correct, make transcription difficult.

Musical scores are not a full set of instructions and a performance is an interpretation of a score and therefore a score is not necessarily an inverse transformation of a performance. Early transcription systems such as [9] used an envelope extractor to determine onsets with the decay time being used to determine if a stroke was damped or not with a 'stroke detector' used to gather more musical information about the strike. Many other transcriptions follow a similar paradigm of detecting the onset of an event, extracting features about that onset and finally classifying the stroke. Correlation techniques using models of drum sounds have proved successful in detecting percussive sounds [10] but the creation of templates is problematic as the template is used to pick out sounds which are not necessarily known a-priori. Models of drum sounds have been developed using band pass and low pass impulse responses for snare and bass drums respectively [11], or using sinusoids plus noise [12] where synthesis is made from information about the onset point. Yoshii [13] used adaptive spectral templates to overcome this problem of variable drum sounds.

Classification of instruments has been done using a variety of methods such as Support Vector Machines and Hidden Markov Models. Results from [14] have show little difference in the performance of methods of classification. Further complexity in transcription arises from the ability to play up to four events simultaneously. Fitzgerald [15] used models to seed PSA for analysis. This work was followed by [16] using non negative spectrogram factorization to help resolve the problem of the changeability in the spectrum of percussive sounds.

## 3. FUTURE RESEARCH AREAS

It has been shown that approaches exist in the three main challenges of percussion analysis from music signals however many of the approaches discussed thus far concentrate their attentions on bass and snare drum locations. In many research areas, it would be desirable to greatly increase the diversity of strokes recognised to include cymbal strikes, bell strikes, rim shots, drum rolls, hit hat sounds, and strokes made with different beater types, indeed full transcription of drum tracks would need to include such strikes as symbolic notations exist for these sounds. A sensor based approach to percussion analysis such as those discussed in [17] could be employed, but only camera based approaches do not alter either the sound of the drum kit or the drummers performance and a signal analysis approach is therefore desirable. A study of different timbres from snare drums [18] shows how differing stroke types or to a degree, musical expression within a percussive performance can begin to be analysed in real-time opening up further possibilities for transcription and score following. While model based approaches would seem appropriate with regard to rim shots and extraction of these has been successful, with cymbal sounds, difficulties arise from their spectral and temporal envelope. The decay time of these is often slow resulting in false or missed triggers from threshold based algorithms. These decays would need to be considered unwanted noise in the identification of strikes occurring during their envelope. Differentiation between cymbal strikes such as bell and standard strikes could be made using methodologies outlined in [18] or [10] where distinct envelope differences exist. Hi-hats offer much variability of sound through the use of the pedal and beaters, and offer further challenges for detection and transcription. Many drum beats use the hi-hat to carry the pulse of the beat and real-time extraction or non real time classification of these stroke types would be a desirable research aim. These improvements to machine listening of percussion offer exciting future possibilities for automated transcription and realtime analysis in many areas.

### 3.1. Machine Musicianship

More information about the rhythmic surface being attended to could improve beat tracker entrainment. It has been stated that beat tracking without prior knowledge of the musical style is probably impossible [19] and this is almost certainly the case in creating generic algorithms for all music, although following music where percussion exists or following a drummer could be improved on. Beat trackers currently focused on downbeat locations which commonly coincide with bass drum or snare drum hits could take for example hi-hat hits as a tactus where the rhythmic surface indicates. Real-time examination of percussive gesture while currently too computationally expensive, could be used in the development of autonomous musical agents, able to react to expressive variations from a human percussionist. Such a system could feed existing performance software allowing sequenced material to be triggered at the correct time at the correct pace or to respond to a percussionist with autonomous improvisation.

### 3.2. MIR Database Applications

Accurate transcription of percussion could lead to automatically populating databases with more content regarding tempo, rhythm, and style leading to more diverse query criteria and improving accuracy. Musicologists could benefit from a more feature rich database or allow for the creation of query by beat-boxing or other spoken input [20] where tracks could be reported from a database based on the rhythmic structure of the percussion.

## 4. MONITORING A STUDENT DRUMMER

By way of conclusion, these research areas converge with possibilities for monitoring or analysing a percussive performance. A system is envisaged capable of monitoring student drummers. Machines could be developed capable of monitoring performance against a pattern which has been set as an exercise, offering remedial exercises where necessary such as slowing the pace down where difficulty is encountered, comparing performance against the notated score and measuring criteria from balance between component parts of the kit to accent and expression. Real time applications such as beat tracking could provide immediate feedback while non real time analysis could be used to review performance indicating to both student and tutor areas which require attention. While such a system should not seek to replace a human tutor capable of analysing posture, grip, and personal style, such a system could be beneficial in the early learning stages where repetitive exercises are conducted and such a project would further research into many of the areas discussed above. By starting from analysis of simple repetitive patterns covering a wider variety of stroke types, research could move forward to more complex arrangements possibly even analysis of improvisation.

## 5. REFERENCES


[1] T.D. Rossing, *Science of Percussion Instruments*, World Scientific Publishing Co Pte Ltd, 2000.

[2] T.M. Gifford and A.R. Brown, "Listening for noise: An approach to percussive onset detection," ACMA, 2008.

[3] J. Bello, L. Daudet, S. Abdallah, C. Duxbury, M. Davies, and M. Sandler, "A Tutorial on Onset Detection in Music Signals," *Speech and Audio Processing, IEEE Transactions on*, vol. 13, 2005, pp. 1047, 1035.

[4] A. Klapuri, "Sound onset detection by applying psychoacoustic knowledge," *Proceedings of the Acoustics, Speech, and Signal Processing, 1999. on 1999 IEEE International Conference - Volume 06*, IEEE Computer Society, 1999, pp. 3089-3092.

[5] N. Collins, "Drumtrack: Beat Induction from an Acoustic Drum Kit with Synchronised Scheduling," *Proceedings of the International Computer Music Conference, Barcelona.* , 2005.

[6] M. Goto and Y. Muraoka, "A beat tracking system for acoustic signals of music," *Proceedings of the second ACM international conference on Multimedia*, San Francisco, California, United States: ACM, 1994, pp. 365-372.

[7] M.S. Puckette, M.S.P. Ucsd, T. Apel, U.(. Edu, D.D. Zicarelli, and C.(.C. Com, "Real-time audio analysis tools for Pd and MSP," *Proceedings, International Computer Music Conference. San Francisco: International Computer Music Association*, 1998, pp. 109-112.

[8] Xavier Rodet Et Al "Improving score to audio alignment: Percussion alignment and Precise Onset Estimation," 2004.

[9] A. Schloss, "On the Automatic Transcription of Percussive Music: From Acoustic Signal to High Level Analysis," 1985.

[10] F. Gouyon, F. Pachet, and O. Delerue, "On the Use of Zero-Crossing Rate for an Application of Classification of Percussive Sounds," *PROCEEDINGS OF THE COST G-6 CONFERENCE ON DIGITAL AUDIO EFFECTS (DAFX-00*, 2000.

[11] A. Zils, F. Pachet, O. Delerue, and F. Gouyon, "Automatic Extraction of Drum Tracks from Polyphonic Music Signals," *Proceedings of the First International Symposium on Cyber Worlds (CW'02)*, IEEE Computer Society, 2002, p. 179.

[12] J. J Sillanpää, A. Klapuri, J. Seppänen, and T. Virtanen, "Recognition of acoustic noise mixtures by combined bottom-up and top-down processing," *proc. European Signal Processing Conference, EUSIPCO 2000* , 2000.

[13] K. Yoshii, M. Goto, and H.G. Okuno, "AdaMast: A Drum Sound Recognizer Based on Adaptation and Matching of Spectrogram Templates," *MATCHING OF SPECTROGRAM TEMPLATES, ISMIR 2004*, 2004, pp. 184--191.

[14] P. Herrera, A. Yeterian, R. Yeterian, and F. Gouyon, "Automatic Classification of Drum Sounds: A Comparison of Feature Selection and Classification Techniques," *PROCEEDINGS OF 2ND INTERNATIONAL CONFERENCE ON MUSIC AND ARTIFICIAL INTELLIGENCE*, 2002, pp. 69--80.

[15] D. Fitzgerald, B. Lawlor, and E. Coyle, "Prior Subspace Analysis for Drum Transcription," *114TH AES CONVENTION AMSTERDAM MARCH 22 ND –25 TH 2003*, 2003.

[16] J. Paulus and T. Virtanen , "Drum transcription with non-negative spectrogram factorisation," *European Signal Processing Conference (EUSIPCO)*, 2005.

[17] A.R. Tindale, A. Kapur, G. Tzanetakis, P. Driessen, and A. Schloss, "A comparison of sensor strategies for capturing percussive gestures," *Proceedings of the 2005 conference on New interfaces for musical expression*, Vancouver, Canada: National University of Singapore, 2005, pp. 200-203.

[18] M.U. Adam Tindale (Music Technology Department, U.O.V. Ajay Kapur (Electrical and Computer Engineering, and M.U. Ichiro Fujinaga (Music Technology Department, "Towards Timbre Recognition of Percussive Sounds," 2004.

[19] N. Collins, "Towards Autonomous Agents for Live Computer Music: Realtime Machine Listening and Interactive Music Systems," University of Cambridge, 2006.

[20] O. Gillet and G. Richard, "Drum loops retrieval from spoken queries," *J. Intell. Inf. Syst.*, vol. 24, 2005, pp. 159-177.