# SCORE-BASED ANALYSIS OF EXPRESSIVE PERFORMANCE

*Simon Dixon*
Austrian Research Institute for Artificial Intelligence, Vienna, Austria

## ABSTRACT

We outline a system for automatic analysis of audio recordings of known musical works, by utilising the musical score to aid the signal processing algorithms. The proposed system matches the audio data to the score note by note, predicts the timing of future notes and then searches in the neighbourhood of the prediction to estimate the actual onset time. In this paper, we address possible signal processing approaches for processing solo piano music, and describe the planned architecture of the rest of the system. We present results of testing the signal processing algorithm on a performance of a Mozart piano sonata. The motivation for this work is the analysis of expressive performance, that is, measuring the subtle interpretative choices which distinguish the great masters of performance.

## 1. BACKGROUND

Automatic analysis of audio has traditionally focussed on speech related data processing rather than music, and the majority of studies related to musical data have used specially chosen or prepared data, in order to ensure certain restrictions such as in the range of instruments and pitches or in the degree of polyphony. On the other hand, performance researchers have either avoided audio altogether, by using special instruments such as the recording pianos produced by Bösendorfer and Yamaha, or they have worked with audio, possibly with the help of analysis tools, but with final measurements being based largely on human judgement. Such a process is laborious and error-prone. Our aim is to produce software which will aid in this task, and enable large-scale analysis of audio recordings of professional performances.

Research on the analysis of musical audio falls into several categories. One of the recurring themes is automatic transcription, systems that take a recording as input and produce a musical score as output. Second, there is music information retrieval, a rapidly growing field specialising in the classification, indexing and retrieval of musical data, mostly for internet, database and library applications. A third area of interest is in real time performance systems, whether for automatic accompaniment of a soloist or small group playing traditional music, or for interactive improvisation, or for synchronisation of devices such as lights, video, animation or recording equipment with music. Finally, another area of interest is the analysis of expressive performance, in which the performer's tempo, dynamic and articulation choices (measured relative to the score) are studied in order to learn more about the practice of music interpretation.

We now briefly review work in these areas as it relates to the present project. Over the last 30 years, many attempts have been made to develop an automatic transcription system, that is, a computer program which produces a musical score directly from audio data, ignoring fine details such as expressive timing and dynamics (e.g. Moorer, 1975; Piszczalski and Galler, 1977; Chafe et al., 1985; Kashino et al., 1995; Martin, 1996; Marolt, 2001; Klapuri, 1998; Sterian, 1999; Klapuri et al., 2000; Dixon, 2000a,b; Raphael, 2002a; Griebel, 2002). Certain features are common to many of these systems: producing a time-frequency representation of the signal, finding peaks in the frequency dimension, tracking these peaks over the time dimension to produce a set of partials, and combining the partials to produce a set of notes. The differences between systems are usually related to the assumptions made about the input signal (for example the number of simultaneous notes, types of instruments, fastest notes, or musical style), and the means of decision making (for example using heuristics, neural nets or probabilistic reasoning).

Closely related to transcription is the work on audio beat tracking (e.g. Desain and Honing, 1989; Large and Kolen, 1994; Goto and Muraoka, 1995, 1999; Scheirer, 1998; Cemgil et al., 2000; Eck, 2000; Dixon, 2001a). Particularly relevant is the onset or event detection parts of these systems, which tend to have a time resolution that is better than that of transcription systems.

Other related projects are automatic accompaniment systems (e.g. Dannenberg and Mukaino, 1988; Raphael, 2001, 2002b) and the score following algorithm of (Pardo and Birmingham, 2002). By aligning the performance with the score at each score event, these systems are implicitly generating a tempo curve, an important part of performance expression. However, due to their real time requirements, it is probable that systems without this restriction would be capable of better results.

The extraction of performance parameters from an audio recording is, in a sense, the inverse of the transcription task. The small asynchronies and variations which are discarded as noise by a transcription system are the performance data which the performance researcher seeks to ascertain. Further, performance analysis usually assumes a known score on which the performance is based, which is the reference point for all measurements.

To our knowledge, the only system to be built which addresses this issue directly is a prototype described by Scheirer (1995), which, given a simple score, attempts to measure onset and offset times and amplitudes of all played notes. Three methods are presented for finding onsets in a monophonic context, based respectively on high frequency energy, RMS power and the output from a comb filter tuned to the partials of the target tone. For onsets in a polyphonic context, only one method is given, similar to the comb filter method but restricted to the fundamental and any partials of the target tone which do not occur in any other simultaneous tone. The release time is calculated as the point at which the energy of a tone drops below 5% of its peak or a new onset (at the same pitch) is detected, but this method is not at all successful. Amplitude is calculated from the log of the peak filter output, which is then scaled linearly to a MIDI value for comparison with the input data.

ISBN 3-931852-67-9
ISSN 1617-6847

## 2. AIMS

The aim of this work is to develop a tool to assist in the automatic analysis of performed music. As such, it forms part of a large project using artificial intelligence techniques to investigate piano performance (Widmer, 2002; Goebl and Dixon, 2001; Dixon et al., 2002). We plan to extend previous work on onset detection (Dixon, 2001c), beat tracking (Dixon, 2001a,b) and automatic transcription (Dixon, 2000a,b) by taking advantage of the known score information, in order to develop a robust performance analysis system for solo piano music. This paper describes the signal processing techniques being considered, shows results for a typical concert work, and concludes with an outline of the planned architecture of the complete system.

## 3. METHOD

Various filtering techniques were developed and tested on a range of data extending from single piano tones to complete performances. The data was obtained from a Bösendorfer SE290 computer monitored grand piano. This enabled precise evaluation of the results, since it provides precise measurements of timing and velocity for all notes. A filtering technique based on the chirp z-transform was used to compute the response of a bank of filters tuned to the fundamental and harmonics of the target tone. The signal was windowed with a 20ms Hanning window, zero-padded to 23.2ms (1024 samples at 44.1kHz sampling rate), with a hop size of 5ms (i.e. 75% overlap).

To detect onsets, the log amplitude of each of the harmonics was measured in a 600ms window around the expected time of the note, and the time of sharpest attack before the peak value is found. This gives a set of onset times for each partial up to the 50th partial or the Nyquist frequency. It was found that the onset time is best estimated by the mean of the onsets of the partials, as opposed to the median and mode values. Results were evaluated by comparison with the known onset times from the Bösendorfer measurements.

## 4. RESULTS AND DISCUSSION

Although the system is still in an early stage of development, results appear to be quite promising. The results in this section are based on testing with Mozart's piano sonata K.332, performed by the Viennese pianist Roland Batik on a Bösendorfer SE290 computer monitored grand piano. Figure 1 shows a histogram of the error in onset detection: of the 9012 notes, 41% are within 20ms of the measured onset and 69% within 40ms, with a mean absolute error of 34ms. The systematic error (bias) is 2ms.

There are many areas in which the results can be improved. One fundamental problem is that the known score information is not as yet being used to guide the system. The information from the score about simultaneous notes is extremely valuable in determining which partials are unique to any note at a particular time. For example, we consider the first two notes of the 2nd movement, which are an octave apart and notated as simultaneous. In the performance, the higher note precedes the lower one by 36ms, so there is no problem finding the onset of the higher tone (see Figure 2). But since the notes share many partials, the system incorrectly finds the onset of the lower note to be about the same as the higher one (see Figure 3, where only the fundamental and 5th partial of the lower tone are clearly seen as coming later). To correct for this, the system should distinguish between the two notes by using the partials which are not common to both tones (i.e. the odd partials of the lower tone).

As well as the problem of interference from simultaneous notes, repeated notes will also cause errors in the current version of the system if they occur within the search window (300ms either side of the target tone). There are numerous cases in the test piece where this occurs (e.g. trills), and the current system makes no attempt to correct for this situation, which could be easily done, for example by reducing the window size in these situations.

The accuracy of signal analysis is strongly dependent on the suitability of the data model. For a restricted class of signals, a more accurate model can be specified, resulting in the possibility of developing more accurate analysis methods. In this work we noted a frequency dependence of measured attack times, specifically that the lower partials have a longer time response than higher frequencies, resulting in a frequency-dependent error in onset detection. Given that we have a large database of single piano tones, it will not be difficult to analyse this data and calculate a proper compensation for this effect. Other features, such as the tuning of the specific piano and the inharmonicity (stretching) of partials, should also be accounted for to achieve more accurate results from the filterbank.

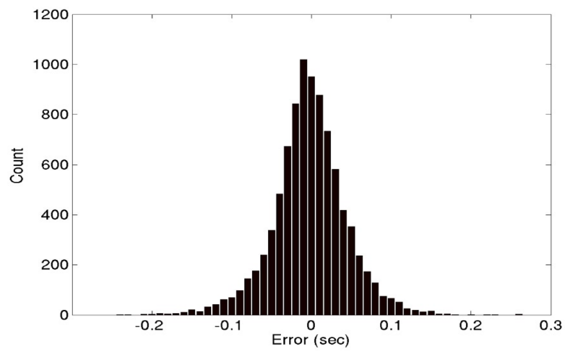


**Figure 1:** Histogram of error in onset detection for Mozart sonata K.332

ISBN 3-931852-67-9
ISSN 1617-6847

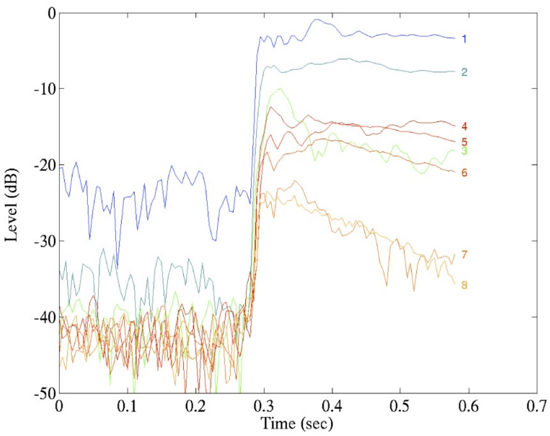


**Figure 2:** Filter output of first 8 partials of the first note of the second movement of sonata K.332.

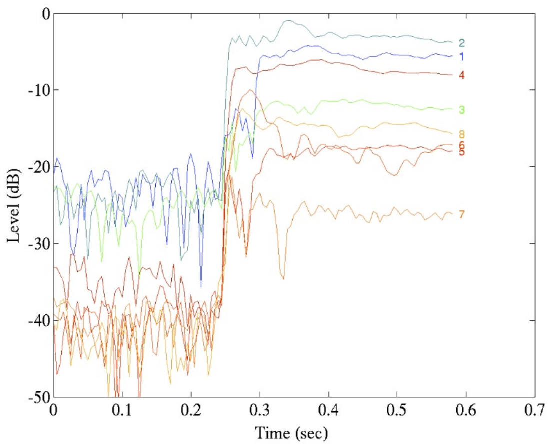


**Figure 3:** Filter output for the first 8 partials of the 2nd note of the 2nd movement of sonata K.332 (see text). The onset of the 2nd note is at 0.3s

## 5. CONCLUSIONS

We have briefly described the signal processing part of a system for analysing audio recordings of musical performances. The complete system will contain a score tracker, using a hidden Markov model or dynamic programming, which tracks the position of the audio signal relative to the score, in order to estimate the expected onset times of notes. This will allow a much narrower search window for finding target tones, and therefore less interference from repeated and harmonically related tones. Removing shared harmonics from the average onset times, and possibly also weighting the harmonics by a certainty factor, is also expected to improve the results considerably.

## 6. ACKNOWLEDGEMENTS

This work was supported by the START programme (project Y99-INF) of the Austrian Federal Ministry of Education, Science and Culture (BMBWK). The Austrian Research Institute for Artificial Intelligence also acknowledges the basic financial support of the BMBWK. Special thanks to the Bösendorfer Company, Vienna, for some of the performance data used in this work.

## 7. REFERENCES

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

2. Chafe, C., Jaffe, D., Kashima, K., Mont-Reynaud, B., & Smith, J. (1985). Techniques for note identification in polyphonic music. In Proceedings of the International Computer Music Conference, San Francisco CA. International Computer Music Association.

3. Dannenberg, R. and Mukaino, H. (1988). New techniques for enhanced quality of computer accompaniment. In Lischka, C. and Fritsch, J., editors, Proceedings of the International Computer Music Conference, pages 243-249, San Francisco CA. International Computer Music Association.

4. Desain, P., & Honing, H. (1989). Quantization of musical time: A connectionist approach. Computer Music Journal, 13(3):56-66.

5. Dixon, S. (2000a). Extraction of musical performance parameters from audio data. In Proceedings of the First IEEE Pacific-Rim Conference on Multimedia, pages 42-45.

6. Dixon, S. (2000b). On the computer recognition of solo piano music. Mikropolyphonie, 6.

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

8. Dixon, S. (2001b). An interactive beat tracking and visualisation system. In Proceedings of the International Computer Music Conference, pages 215-218, San Francisco CA. International Computer Music Association.

9. Dixon, S. (2001c). Learning to detect onsets of acoustic piano tones. In Proceedings of the Workshop on Current Directions in Computer Music Research, pages 147-151, Barcelona, Spain. Audiovisual Institute, Pompeu Fabra University.

ISBN 3-931852-67-9
ISSN 1617-6847

10. Dixon, S., Goebl, W., & Widmer, G. (2002). Real time tracking and visualisation of musical expression. In Music and Artificial Intelligence: Second International Conference, ICMAI2002, pages 58-68, Edinburgh, Scotland. Springer.

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

12. Goebl, W., & Dixon, S. (2001). Analysis of tempo classes in performances of Mozart sonatas. In Proceedings of VII International Symposium on Systematic and Comparative Musicology and III International Conference on Cognitive Musicology, pages 65-76, University of Jyväskylä, Finland.

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

14. Goto, M., & Muraoka, Y. (1999). Real-time beat tracking for drumless audio signals. Speech Communication, 27(3-4):331-335.

15. Griebel, H. (2002). Time-Frequency Methods for Pitch Detection. PhD thesis, Technical University of Vienna, Institute for Analysis and Technical Mathematics.

16. Kashino, K., Nakadai, K., Kinoshita, T., & Tanaka, H. (1995). Organization of hierarchical perceptual sounds: Music scene analysis with autonomous processing modules and a quantitative information integration mechanism. In Proceedings of the International Joint Conference on Artificial Intelligence.

17. Klapuri, A. (1998). Automatic transcription of music. Master's thesis, Tampere University of Technology, Department of Information Technology.

18. Klapuri, A., Virtanen, T., & Holm, J.-M. (2000). Robust multipitch estimation for the analysis and manipulation of polyphonic musical signals. In Proceedings of the COST-G6 Conference on Digital Audio Effects, Verona, Italy.

19. Large, E., & Kolen, J. (1994). Resonance and the perception of musical meter. Connection Science, 6:177-208.

20. Marolt, M. (2001). SONIC: Transcription of polyphonic piano music with neural networks. In Proceedings of the Workshop on Current Directions in Computer Music Research, pages 217-224, Barcelona, Spain. Audiovisual Institute, Pompeu Fabra University.

21. Martin, K. (1996). A blackboard system for automatic transcription of simple polyphonic music. Technical Report 385, Massachussets Institute of Technology Media Laboratory, Perceptual Computing Section.

22. Moorer, J. (1975). On the Segmentation and Analysis of Continuous Musical Sound by Digital Computer. PhD thesis, Stanford University, CCRMA.

23. Pardo, B. and Birmingham, W. (2002). Improved score following for acoustic performances. In Proceedings of the International Computer Music Conference, pages 262-265, San Francisco CA. International Computer Music Association.

24. Piszczalski, M., & Galler, B. (1977). Automatic music transcription. Computer Music Journal, 1(4):24-31.

25. Raphael, C. (2001). Synthesizing musical accompaniments with Bayesian belief networks. Journal of New Music Research, 30(1):59-67.

26. Raphael, C. (2002a). Automatic transcription of piano music. In Proceedings of the 3rd International Conference on Musical Information Retrieval.

27. Raphael, C. (2002b). A Bayesian network for real-time musical accompaniment. In Dietterich, T. G., Becker, S., and Ghahramani, Z., editors, Advances in Neural Information Processing Systems 14, pages 1433-1439, Cambridge MA. MIT Press.

28. Scheirer, E. (1995). Extracting expressive performance information from recorded music. Master's thesis, Massachusetts Institute of Technology, Media Laboratory.

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

30. Sterian, A. (1999). Model-Based Segmentation of Time-Frequency Images for Musical Transcription. PhD thesis, University of Michigan, Department of Electrical Engineering.

31. Widmer, G. (2002). In search of the Horowitz factor: Interim report on a musical discovery project. In Proceedings of the 5th International Conference on Discovery Science, Berlin. Springer.

ISBN 3-931852-67-9
ISSN 1617-6847