EXTRACTION OF PERFORMANCE PARAMETERS FROM ACOUSTIC RECORDINGS OF PIANO MUSIC

A. Earis, D.J. Tait and B.M.G. Cheetham
University of Manchester, Oxford Road, Manchester, M13 9PL
e-mail: earisa@cs.man.ac.uk

1. ABSTRACT

This paper concerns the characterisation of expressive musical performance in acoustic recordings of piano music. The main focus is the automatic extraction of timing and dynamics. Measurement of timing requires an estimation of the inter-onset times of individual notes. The characterisation of dynamics involves the evaluation of note volume both in a melody and within individual chords. The variables are estimated with reference to a machine readable version of the musical score. Synchronisation of score and recording is accomplished by correlating the music signal with piano waveforms synthesised from the score, using the discrete Fourier transform in sixteen sub-bands over the complete frequency range of the piano keyboard and with frequency-dependent window sizes (50 to 400ms). The paper highlights a number of problems which concern the measurement of performance dynamics i.e. interference between the harmonics of simultaneous tones, and the complex nature of piano tone itself. Applications for both performers and musicologists are discussed. The parameters may be used to augment a musical score to illustrate the differences between individual performances.

2. INTRODUCTION

Digital Signal Processing (DSP) techniques may be used to resolve musical sounds in both time and frequency. It is thus possible to analyse acoustic recordings to automatically extract expressive performance information, with reference to a digitised score of the work being performed. This allows the different expressive features to be measured and parameterised.

Features of expressive musical performance that we may wish to analyse include timing (note onset and offset times), dynamics (variations in the intensity of notes and chords), tone quality (timbre), texture, articulation, intonation and pitch. The nature and number of these features depend on the instruments being played and their characteristics. For a piano, the performer controls timing, dynamics and articulation. A violinist, in addition, can vary pitch and a flautist can influence the tone quality (timbre) of the note being played. The main focus of this paper is the extraction and parameterization of timing and dynamics from performances of piano music.

Timing is the most fundamental aspect of musical performance. The performer can use variations in the attack and release of notes and chords to shape musical structure and phrasing. Expressive timing variations can be very subtle, the human ear being capable of detecting differences of the order of tens of milliseconds [1]. It is therefore necessary for very accurate timing measurements to be made. Variations in intensity, which can be sudden or gradual, are used by performers in many different ways to characterise the music.

Much has been learned about performance characterisation by studying recordings made on MIDI (Musical Instrument Digital Interface) instruments. However, the information thus recorded is limited to note onset, note offset and volume [e.g. 2,3,4,5]. This study involves the extraction of performance information from acoustic recordings of piano music. Acoustic recordings, which contain a much more complete range of expressive performance features than MIDI, are universally available providing a vast supply of invaluable musical information. Recordings spanning over a century are available, thus enabling the analysis of a wide range of performance periods and styles.

There have been a number of studies of acoustic recordings, including the analysis of timing patterns [e.g. 2,6] and dynamics [e.g. 7]. Such studies are generally limited to piano performances. In the papers cited, the measurements were made by manually studying the waveforms of the sound recording and locating note onsets using both visual and auditory clues. Dynamics measurements were restricted to the intensities of successive tones or groups of tones or chords, rather than individual notes within chords. A method of automatically extracting the data will be of considerable advantage in the collection and analysis of a large amount of performance information, thus removing the restriction of having to concentrate on a small number of notes and/or performers.

3. SYNCHRONISATION OF SCORE AND ACOUSTIC RECORDING

In order to characterise an acoustic recording, individual notes and chords must be identified and related to the musical score. This has been achieved by comparing a computer-generated model of the frequency spectrum of each 'event' in a musical score with successive time segments of a spectrograph of the acoustic recording over a suitable time-span. The musical score is stored digitally as a MIDI file. An event is described as the onset, or attack, of a note or chord as would be synthesised by a MIDI piano. Each event is characterised by a harmonic spectrum i.e. one or more fundamental frequencies, their amplitudes and the amplitudes of the harmonics. The frequency spectrum for each synthesised note is based on a very simple model of piano tone where the amplitudes of the first five harmonics are in the ratio 5:4:3:2:1. The spectrograph of the acoustic recording is calculated using the discrete Fourier Transform (DFT) performed at intervals of 0.01 seconds with 95% overlapping windows. The DFT uses variable window sizes (50ms to 400ms), dependent on the lowest frequency component in the event.

Agreement between an event spectrum and a spectrograph segment is measured by computing the cross-correlation between them. The highest value of cross-correlation obtained over the chosen time-span indicates the most likely time of occurrence within the acoustic recording of the sound corresponding to the event indicated in the score. The result of performing such an analysis for each event within the score is an array of timing measurements which augments the MIDI file with the precise onset times of each note or chord as located within the acoustic recording. These timing measurements can be used to synthesise music from a MIDI recording with the exact timing patterns of a particular acoustic recording. Furthermore, the musical score could be annotated with the extracted timing information.

Fig. 1 shows the cross-correlation for four different notes for a C major scale in thirds. The locations of individual notes in the acoustic recording can be seen.

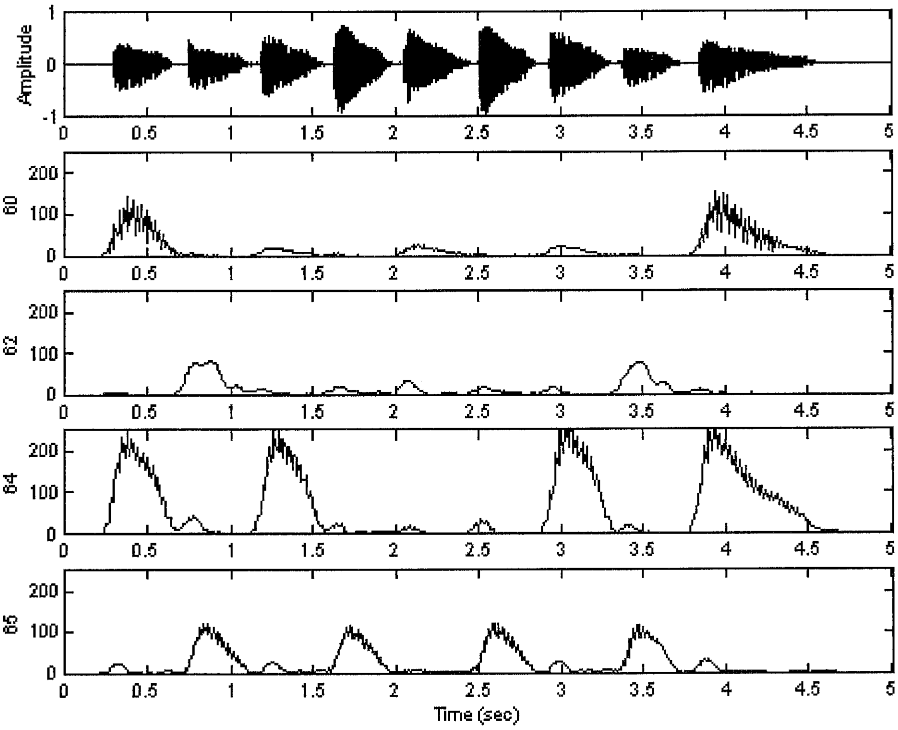

Fig. 1. Correlation between the musical signal and models of the maximum amplitude of each note. The top graph shows the musical waveform and the lower four graphs give the correlations for individual notes (middle C = 60, D = 62, E = 64, F = 65).

In order to test the timing extraction method, timing onset measurements for a musical recording were made in two ways: by the proposed method and by a simple examination of the musical waveform, using visual and aural clues. This showed that the computer generated results were highly accurate, the maximum deviation between the computer measured and manual results being 5 ms. Fig. 2 shows the inter-onset timing, as extracted by the proposed method, for two performances of the opening bars of Beethoven's Moonlight piano sonata by two famous pianists (Moura Lympany and William Kempff). The differences between the two performances can be clearly seen.

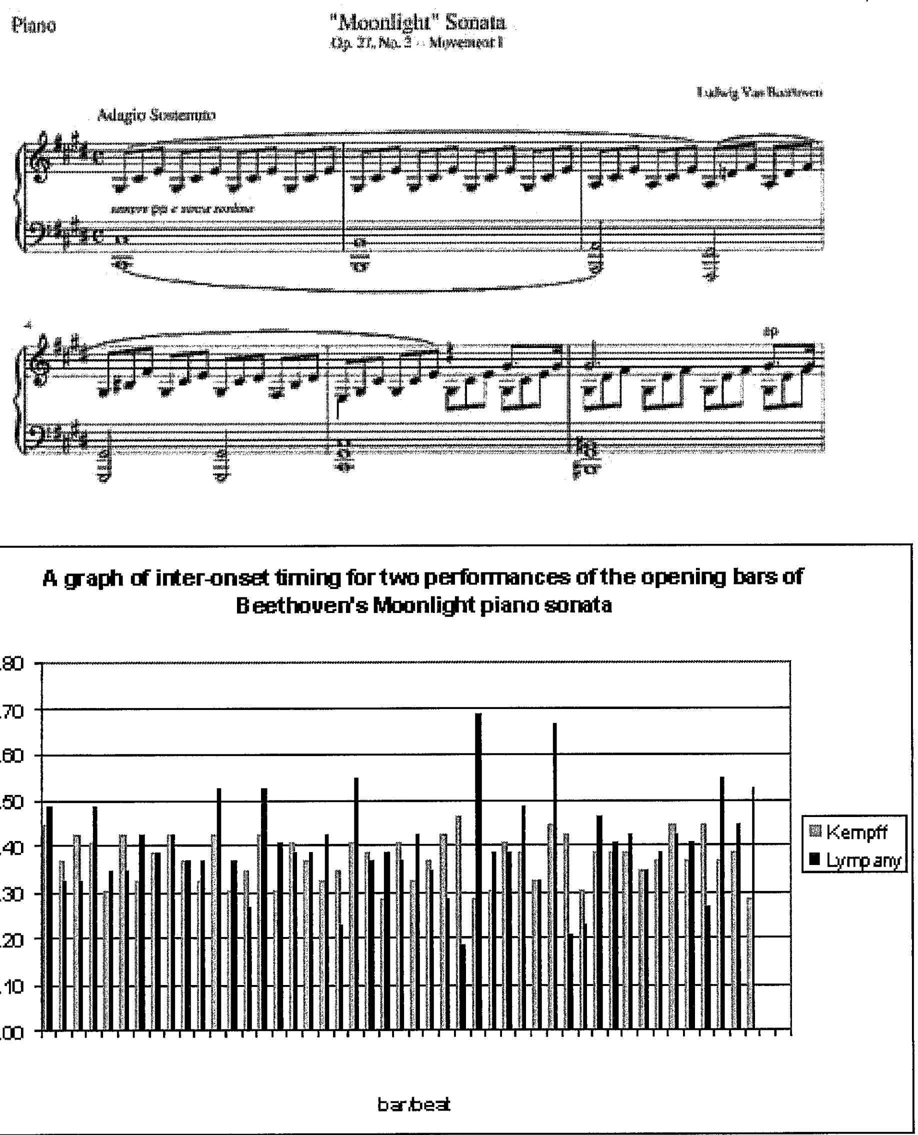

Fig. 2. A graph of inter-onset timing for two performances of the opening bars of Beethoven's 'Moonlight' piano sonata. The score is shown above.

4. MEASURING DYNAMICS

Having identified the locations of events in an acoustic recording, it is straightforward to measure the energy, or volume, of each attack by measuring the peak energy of the waveform at the point of attack. Fig. 3 shows the event energies, or dynamics, for the Lympany recording of the Moonlight piano sonata.

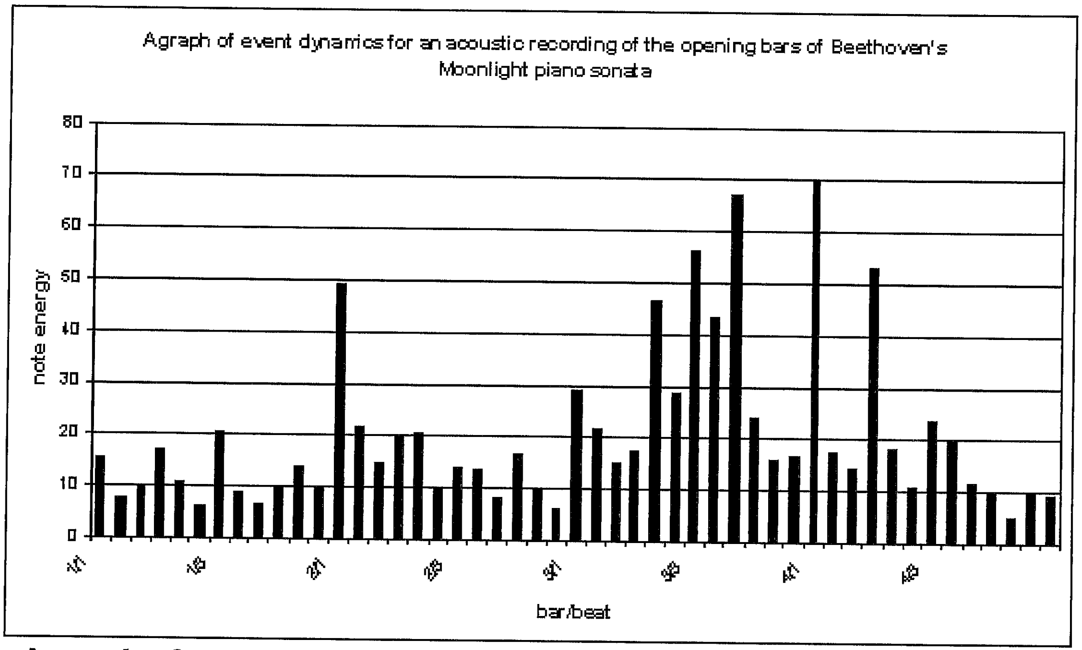

Fig. 3. A graph of event energy, or dynamics, for the opening bars of the Lympany recording of Beethoven's 'Moonlight' piano sonata.

However, if the objective is to measure the volume of individual notes to study the relative volume of the left hand and right hand, for example, or the melody and accompaniment, the problem becomes more complicated. When estimating the relative intensities of simultaneous tones, or individual notes within chords, overlapping harmonics and unpredictable phase relationships between the harmonics affect the overall amplitude of each note [8]. This problem is illustrated in Fig. 4. This is an analysis of the same opening bars of Beethoven's Moonlight piano sonata, but for a computer generated performance using synthesised piano tones. Each line (treble, tenor, bass) is recorded separately and the maximum amplitude of each note is displayed in the three graphs. This analysis is then performed on a recording of all lines played simultaneously, i.e. a complete performance. The results show that both constructive and destructive interference between harmonics does occur.

Fig. 4. Analysis of the individual lines of the opening of Beethoven's Moonlight piano sonata of a computer-generated recording using synthesised piano tone (top left – right hand triplets; top right – tenor; bottom left – bass). The graph (bottom right) is an analysis of the complete performance with all lines played simultaneously.

In addition, piano tone is much more complex than our simple model suggests. Fig. 5 shows the fractional amplitude of the first five harmonics of each note of a C major chromatic scale, starting at middle C and rising one octave, as measured from a recording of an acoustic piano and calculated using the DFT. There is apparently random variation in the relative strengths of the individual harmonics. This variation is due to a number of factors concerning the physical design of the piano, the non-linear frequency response of the soundboard and sound reflections both within the case and within the room.

Fig. 5. Fractional amplitude of the first five harmonics of each note of a C major scale, averaged over several recordings.

A simple way of estimating individual note dynamics is to sum the first the amplitudes of the five harmonics of a given note at its point of attack (as indicated by the event time found by the correlator). However, this has major limitations due to the factors described above. Fig. 6 shows the comparison of measurements of dynamics of the right hand triplet figuration of the opening bars of the Moonlight piano sonata for the two pianists.

Fig. 6. A graph of triplet dynamics for two performances of the opening bars of Beethoven's 'Moonlight' piano sonata.

5. APPLICATIONS

An understanding of the ways in which performers achieve musical expression, and the 'fingerprints' of different performance styles, is essential for performers and musicologists. A complete understanding of the different ways of performing a work would be of great benefit to a performer in forming their own interpretation of the work.

Whilst the graphical format in Figs. 2 and 3 is satisfactory for a small number of results, the complete analysis of a piece, with complex polyphony and differing note lengths would generate a substantial amount of data. A 'third dimension' may be added to a musical score to indicate the expressive performance information. For example, colour may be added to a musical score to provide a clear visual way of demonstrating an individual's performance dynamics. The addition of timing information requires a scheme for indicating whether notes are earlier or later than the underlying beat, and to what degree this is the case. Possible ways of indicating this information include tilting the note stems to indicate whether a note is earlier or later, changing the note size, or adding a feature above the staves to indicate whether the music speeds up or slows down at that point.

6. CONCLUSIONS

This paper discusses a method of automatically extracting the expressive timing and dynamics information from an acoustic recording of a musical performance, with reference to a digitised score of the work being performed. Comparisons between timing measurements calculated automatically using this method and those made manually show that the system has a high degree of accuracy. A simple method of measuring performance dynamics is outlined, whereby the dynamics of successive events (notes or chords) are estimated from the peak energy of each event. However, the measurement of individual note dynamics in an acoustic recording poses a number of problems involving interference between the harmonics of simultaneous tones, and the complex nature of piano tone itself.

7. REFERENCES

[1] Sloboda, J. A. (2000). Individual differences in musical performance. Trends in Cognitive Sciences **4**, 397-403.

[2] Repp, B.H. (1990). Patterns of expressive timing in performances of a Beethoven minuet by nineteen famous pianists. Journal of the Acoustical Society of America **88**, 622-650

[3] Repp, B.H. (1996). The dynamics of expressive piano performance: Schumann's "Traumerei" revisited. Journal of the Acoustical Society of America **100**, 641-650.

[4] Windsor, W.L. & Clarke, E.F. (1997). Expressive timing and dynamics in real and artificial musical performance: using an algorithm as an analytical tool. Music Perception **15**, 127-152.

[5] Juslin, P.N. & Madison, G. (1999). The role of timing patterns in recognition of emotional expression from musical performance. Music Perception **17**, 197-221.

[6] Repp, B.H. (1998). A microcosm of musical expression: I. Quantitative analysis of pianists' timing in the initial measures of Chopin's Etude in E major. Journal of the Acoustical Society of America **104**, 1085-1100.
[7] Repp, B.H. (1999). A microcosm of musical expression: II. Quantitative analysis of pianists' dynamics in the initial measures of Chopin's Etude in E major. Journal of the Acoustical Society of America **105**, 1972-1988.
[8] Repp, B.H. (1993). Some empirical observations on sound level properties of recorded piano tones. Journal of the Acoustical Society of America **93**, 1136-1144.