

AUTOMATED QUANTISATION AND TRANSCRIPTION OF ORNAMENTS FROM AUDIO RECORDINGS

Georg Boenn

University of Glamorgan

Atrium, Department of Music and Drama, Division of Music and Sound

ABSTRACT

We propose a new method for rhythm quantisation and measurement of expressive timing. This paper focuses on the automatic quantisation and rhythmic transcription of syncopated rhythms and baroque ornaments, e.g. *apoggiaturas*, *mordants* and *trills* from time-tagged audio recordings without knowing the score in advance. We demonstrate the transcription of the Aria of J. S. Bach's Goldberg Variations, BWV 988, recorded by Glenn Gould in 1955 and 1981¹.

1. INTRODUCTION

1.1. Motivation

Tools for Automated Transcription are in demand by composers, musicologists, musicians, improvisers, publishing houses, archives and libraries, facilitating also the creation and use of online databases, data-mining and possible integration into real-time performances. We hope that the research described here will also be helpful in the area of music cognition.

Research on musical timing in performances helps us to gain better understanding of musical performance practices. The application of sound engineering and a century of recording history provide rich grounds for research in this area. It helps us to investigate the musical motivations for certain *agogics* or manners of playing. One important outcome is that detailed characteristics of performances can be established for specific musical styles and even for certain performers. Changes and shifts in aesthetics can be traced and identified more easily and more accurately than before on the basis of historic recordings, as is done for example in the Mazurka Project at CHARM².

1.2. Problem statement

We want to grasp the subtleties in rhythmic expression that cover the use of *agogics* (*rubato*, expressive timing), as well as the use of ornaments like the *trill* or *mordant*, which are to be played in relation to the musical context, neighbouring durations and the current tempo, but also offer room for tempo fluctuations and deviation from

straightforward metrical playing. Certain manners of playing, which are not directly encoded in notation, can become characteristic of a musical style. The playing of over-dotted notes in baroque music is a much debated example of such a practice [6][7]. Finally, the fact that no musician plays metronomically exactly stems from the human need for musical expression, which can only materialise itself in time. Time-tagged data can be derived from audio or midi files or modelled by a performance editor³. A rhythmic analysis of timing data has to take into account that the tempo within a musical performance is in a state of constant flux. The two main problems are to detect at the same time tempo changes on different scales⁴ and to identify the metrical relationships between all note-onsets [10].

Musical ornaments are a particularly challenging problem. Trills are traditionally not written out note by note, their execution and timing leaving considerable room for interpretation and expression on the part of the musician. The rapid interchange of two notes might also lead to the perception of a single sound event, where the individual notes fuse together creating an impression of unity, rather than a sequence of discrete note events. On the other hand, trills can also be played out slowly, especially within a calm and slow movement⁵. The challenge for any automated transcription system is that it should distinguish the ornament from other events and correctly identify and transcribe its onset locations. Trills appear in a great variety of stylistic flavours. In order to give some rules to the performers, many educational works of the late baroque and early classical eras indicate with metrical precision how a specific ornament should be executed [6]. However, scholarly works on ornamentations are not consistent [7]. More importantly, many players choose to execute ornaments freely in an improvised, spontaneous manner⁶. Therefore, a transcription algorithm cannot rely simply on pattern matching based upon an encoded metrical table of all 'historically approved' practices. In this paper we will focus on precisely this challenge, which is

³ e.g. Director musices by KTH

⁴ Honing [9] suggests that there are two main timing procedures, one related to global tempo changes like *rubato* playing and the other being independent from the global tempo but relating only to the onset location of the current beat, e.g. *swing* or *grace notes*

⁵ see Tureck, cited in Bazzana[1], p. 232

⁶ "[Hans] Klotz re-affirmed that the metre of ornaments was basically free, the notation only indicative (Klotz, 1984, p.37)."[7]

¹ discography at <http://sonybmgmasterworks.com/artists/glenn Gould/>

² <http://mazurka.org.uk>

the rhythmic transcription of quantised ornaments and onset data collected from audio recordings. We do this first by investigating closely every single event in its rhythmic context and we will finally achieve the kind of notation that is added to critical editions explaining the execution of an ornament⁷. Pitch information and knowledge about the rules of counterpoint have to be added at a later stage, enabling the program to provide common practice shorthand notations.

1.3. Research in this area

We are aware of previous research carried out in the areas of rhythm quantisation and tempo tracking [4], beat tracking [5][8] and automated rhythm transcription [10]. Although none of them addresses the problems arising from ornamentation practices, their general use for transcription of time-tagged audio data is acknowledged. Their models incorporate Bayesian statistics, hidden Markov models, Kalman filtering, neural networks and dynamic programming. The extraction of piano trills has been addressed in [2]. Their method uses a statistical analysis of spectral data from a database of trills only. There has been a recent approach to the detection of chord spreads (arpeggios) and trills within the MPEG-7 context [3]. In the latter it is assumed that ornamentations are always relatively fast sequences of notes, but this assumption is far too general to become a model for all trills and arpeggios no matter which musical context, let alone other kinds of musical ornaments. They are not necessarily played ‘as fast as possible’. On the contrary, if those more slowly improvised events can easily be confounded with other note durations written in the score, it is likely that a neural network trained with data from one pianist has to deal with unexpected events, when confronted with the ornamentation practice of another performer, which might lead to errors in the transcription.

We therefore propose a model that involves no machine learning or neural networking techniques. More importantly, our model does not only look at the immediate predecessor of an event as in the published methods using directed acyclic graphs [4][10], but it rather takes into account each and every connection, forwards and backwards in time, of a single event with all remaining note events within a given analysis frame. In this respect our approach is closer to Dixon’s clustering method used for beat tracking [5].

2. MODEL

2.1. Search for rhythmic groupings

The program takes a list of event onset values from time-tagged audio data. An event can be a musical note, a chord or a silence. The time-tagged data is segregated into single bars regarded as the analysis frame, which represents

⁷ Gould studied the Kirkpatrick edition (1937)[1]. Although this edition contains a transcription of the Aria, Gould does not adhere to it by the letter.

0.018515, 0.019097, 0.035479, 0.019194, 0.020260, 0.047208, 0.085014, 0.097906, 0.092284, 0.106921

Table 1. IOT list of Gould’s 1981 recording, bar 4.

#0	0.0185150001
	0.019097000
	0.0191939995
	0.0202600006
#1	0.0354790017
	0.04720
#2	0.0850140005
	0.0979060009
	0.092284001
	0.106921002

Table 2. Duration classes in bar 4

in effect the prior knowledge of the downbeat’s onset position. The data is then converted into inter-onset-times (IOTs) and normalised. The program calculates for each single event all time differences with the bar’s remaining events. All events whose differences in duration fall below a certain threshold are copied into a common duration class. The threshold is defined as the current normalised event duration d multiplied by a heuristic variable x , where x can vary between 0.15 and 0.5 depending on the value of d . Duplications of lists of a specific duration class are later removed. Remaining lists which have at least two event pointers in common are merged together. Our tests have shown that this process leads to a collection of duration classes that accurately represents, for each bar alone, all events that are meant to have the same duration in the score but whose actual performance value can be different within certain limits. The facts that all differences between all events are taken into account, and that the merging process generates unique lists of duration classes, both ensure that gradual tempo variations can be followed correctly. The following example, see table 1, should illustrate this procedure. One clearly observes the *ritardando* at the end of the bar that prepares the next phrase of the theme⁸. The melody in beat 1 with its two downwards slides is one example of Bach’s ‘written-out’ ornamentations [7]. Gould also plays the *appoggiatura* on beat 2 as a quaver⁹. The grouping algorithm then reveals three duration classes, see table 2. Duration class #0 contains the demisemiquavers of beat 1, #1 holds the semiquavers and #2 represents the quavers of the combined rhythm between left and right hand (beats 2 and 3).

2.2. Search for true note durations

What follows after obtaining the duration classes is the estimation of their correct score durations and onset values. This process first takes the mean value for each duration

⁸ The score is available online at ISMLP, <http://imslp.org>

⁹ Following the Kirkpatrick edition (1937)

#0	0.0192665011	=	10/519
#1	0.0413435027	=	16/387
#2	0.095531255	=	47/492

Table 3. Mean values for each duration class in bar 4

class. Then it searches in the vicinity of the mean for ratios that are simple enough for western common practice notation (CPN). The reason for this approach is based on the fact that CPN uses simple integer ratios to represent note or silence durations as well as a system of time signatures and metrical structures which is based on the same ratios. An algorithm found by Klarenz Barlow for his composition system *Autobus* was adopted by the author and is used to measure the complexity of an integer ratio and subsequently to estimate its usefulness for CPN. Barlow's indigestibility¹⁰ function is a measure for the decomposition of an integer into its prime factors. For example, this measure sorts the natural numbers 1 to 16 into the series: 1, 2, 4, 3, 8, 6, 16, 12, 9, 5, 10, 15, 7, 14, 11, 13. For the quantisation it is assumed that only those durations whose integer ratios have a denominator with a low indigestibility index can be useful for CPN, otherwise the notation would be overly complex. We therefore present the quantiser with lists of precompiled smooth ratios in the interval [0..1] whose denominator has a low indigestibility index. This list is generated by taking the Farey sequence $F(200)$ and filtering out all ratios whose denominators are not b-smooth. For the Gould transcriptions we use lists of ratios with 3- and 5-smooth denominators. The program then searches this list around each mean taken from the duration classes in both directions. The search has to obey two rules, namely to rank the smooth ratios according to their distance from the mean and, more importantly, to weight the smooth ratios according to their indigestibility index measured from the denominator. This is achieved by calculating the indigestibility indices first and then weighting their values by a Gaussian that is centred around the mean of the duration class in question. The first four smooth ratios with the highest weight are fetched from the ratio list because our tests have shown that one of the four is always the 'winner'. The width of the Gaussian, which is normally at 50% of the mean of one duration class, is increased to 75% if the duration class contains only one element. Table 3 shows the mean values for each duration class in bar 4. Table 4 shows which smooth ratios were subsequently found for each duration class. The 'winners' are #0: 1/48, #1: 1/24, #2: 1/12.

2.3. Transcription

The final stage of the quantisation is then to estimate the original score notation by comparing all possible scores based on combinations of the found smooth ratios and to

¹⁰ Barlow uses this function to calculate the optimum intervals of a tuning system. See Barlow's "Musiquantenlehre", course materials obtained by the author, Musikhochschule Cologne, 1991, and re-affirmed through personal communication with Barlow on 28.01.2007

class	ratio	decimal	indigest.	weight
#0	1/48	0.02083	6.67	0.146094382
	1/64	0.01562	6	0.144295514
	3/128	0.02344	7	0.118486807
	1/54	0.01859	9	0.110432364
#1	1/24	0.04167	5.67	0.176427335
	1/32	0.03125	5	0.15742746
	3/64	0.04687	6	0.155107319
	5/128	0.03906	7	0.141090661
#2	1/8	0.125	3	0.227954671
	1/12	0.08333	4.67	0.20065555
	3/32	0.09375	5	0.199693859
	1/9	0.11111	5.33	0.168506429

Table 4. Ratios found for duration classes in bar 4

choose the preferred score notation. This process applies a different function found by Barlow, namely the measurement of the harmonicity of a musical interval. This function makes use of the indigestibility function applied to both elements of a ratio. If one builds the sum of all harmonicities for each smooth ratio of a possible score representation, then this index can be used to sort the list of scores and to search for the optimum score. Finally, the best scores are being rendered using *lilypond*. However, the complexity of a rhythm notation also depends on the underlying meter and because the meter is not known to the quantiser in advance, it has to be estimated as well. The search for a suitable meter in connection with the quantised rhythm of events is based again on Barlow's harmonicity function. The meter that leads to the least complex notation of a rhythmic sequence is preferred.

3. EXPERIMENTS

3.1. The Bach Aria

The Bach Aria was chosen because it is a highly ornamented piece. There are two famous recordings by Glenn Gould that are completely different in style and tempo. The tempo of the latest recording of the Aria is about two times slower than in 1955. In addition, the Aria is repeated each time at the end of the cycle. Besides its rich ornamentations the Aria shows interesting balances between repetition and variation of rhythmic and melodic figures (*Gestalten*).

The time-tags were generated with the C program library *aubio*¹¹ and have been edited manually with *audacity*. A single list of onset data has been produced that represents the combined rhythm of both hands. Gould, like other Keyboard players, sometimes performs inequalities [7] between both hands in order to accentuate a musical event. Thus we were able to explore the effect of independent part playing on our algorithm.

¹¹ see Brossier, P.M., <http://aubio.piem.org>

Recording details:		
year	1955	1981
onsets played (both hands)	359	351
ornaments played	48	49
Series #1 - prior knowledge of downbeat locations:		
onsets quantised correctly	219 (61%)	308 (88%)
ornaments transcribed	31 (65%)	41 (84%)
Series #2 - prior knowledge of single beat locations:		
onsets quantised correctly	352 (98%)	351 (100%)
ornaments transcribed	46 (95%)	49 (100%)
number of critical bars	8	3

Table 5. Quantisation of the Bach Aria

3.2. Results

The kind of ornaments used by Bach are in order of their appearance: Mordant, Appoggiatura, Double Cadence, Cadence, Trill, Arpeggio, Accent and Trill, and Slide. Gould adds in both recordings two inequalities in the alto voice (bars 16 and 24 on beat 2) to create additional 4-3 suspensions. In 1981 he adds an inequality in the bass line in bar 19, beat 3, in order to accentuate the dissonance with the trill of the alto voice. He also resolves with inequality the appoggiatura of the soprano in bar 26.

Table 5 lists the results of the quantisation¹². The optimisation in series #2 shown in table 5 was achieved by recalculating only the bars that have failed in series #1. Those are the bars that were split into single beats before quantisation: bars 17, 19 and 32 for 1981, bars 7, 8, 11, 13, 17, 21, 22 and 32 for 1955.

The ritardando at the end of the first part was transcribed successfully for both recordings (series #1). At the end of the aria (played without repetitions) the last note with appoggiatura is held with fermata in both recordings, however, this bar has been transcribed successfully in series #2 for both recordings.

4. DISCUSSION AND CONCLUSION

The method we presented is successfully able to quantise onsets from performance data extracted from audio recordings, and to transcribe the results into common practice notation. The ornaments and combined rhythms of right and left hands were successfully transcribed from two different piano recordings of Bach's Aria of the Goldberg Variations.

Apart from the downbeat locations, the program has no prior knowledge of the score. The onset calculation is based on an analysis window that hops over the onset data and is aligned to the exact length of one bar. In some cases the size of the window needs to be reduced to the length of a beat. Therefore, the program would profit from a robust beat detection algorithm in order to eliminate the need to determine the window size by hand, although that would

be a useful feature in a more interactive quantisation process, for example when composers work with CAC applications. We will look into clustering of IOTs and statistical analysis in order to add a beat tracker in the future. We would also like to carry out tests in many more different musical styles that feature improvised ornamentation, like for example the virginalists, south-Indian carnatic music and Jazz. Finally, we will evaluate the use of the program within a creative context, for example in computer assisted composition.

5. REFERENCES

- [1] Bazzana, K. *Glenn Gould: the performer in the work : a study in performance practice*. Oxford, 1997.
- [2] Brown, J. C. and Smaragdis, P. "Independent component analysis for automatic note extraction from musical trills", *Journal of the Acoustical Society of America*, 2004.
- [3] Casey, M. and Crawford, T. "Automatic Location and Measurement of Ornaments in Audio Recordings", *Proceedings of the 5th International Symposium of Music Information Retrieval (ISMIR)*, Barcelona, 2004.
- [4] Cemgil, and Kappen, "Monte Carlo Methods for Tempo Tracking and Rhythm Quantisation", *Journal of Artificial Intelligence Research*, 18:45-81, 2003.
- [5] Dixon, S. "Automatic Extraction of Tempo and Beat from Expressive Performances", *Journal of New Music Research* 30:1, pp. 39-58, 2001.
- [6] Efrati, R. R. *Treatise on the execution and interpretation of the sonatas and partitas for solo violin an the suites for cello solo by Johann Sebastian Bach*. Atlantis, Zurich, 1979
- [7] Fabian, A. *Bach performance practice, 1945-1975*. Aldershot, Hampshire, England, 2003.
- [8] Hainsworth, S. and Macloed, M. "Beat Tracking with Particle Filtering Algorithms", *IEEE Workshop on Applications of Signal Processing to Audio and Acoustics*, Mohonk, NY., 2003
- [9] Honing, H. "From Time to Time: The Representation of Timing and Tempo", *Computer Music Journal*, 25:3, pp. 50-61, Boston, Fall 2001
- [10] Raphael, C. "Automated Rhythm Transcription", *Proceedings of the 2nd International Symposium of Music Information Retrieval (ISMIR)*, pp. 99-107, 2001

¹² Results are also available online at <http://dream.cs.bath.ac.uk/transcriptions/>