A DYNAMIC PROGRAMMING APPROACH TO THE EXTRACTION OF PHRASE BOUNDARIES FROM TEMPO VARIATIONS IN EXPRESSIVE PERFORMANCES

Ching-Hua Chuan* **and Elaine Chew†**
University of Southern California Viterbi School of Engineering
*Department of Computer Science and †Epstein Department of Industrial and Systems Engineering
Integrated Media Systems Center, Los Angeles, CA
{chinghuc,echew}@usc.edu

ABSTRACT

We present an approach to phrase segmentation that starts with an expressive music performance. Previous research has shown that phrases are delineated by tempo speedups and slowdowns. We propose a dynamic programming algorithm for extracting phrases from tempo information. We test two hypotheses for modeling phrase tempo shapes: a quadratic model, and a spline curve. We test the two models on phrase extraction from performances of entire classical romantic pieces namely, Chopin's *Preludes Nos. 1* and *7*. The algorithms determined 21 of the 26 phrase boundaries correctly from Arthur Rubinstein's and Evgeny Kissin's performances. We observe that not all tempo slowdowns signify a boundary (some are agogic accents), and multiple levels of phrasing strategies should be considered for detailed interpretation analyses.

1 INTRODUCTION

Musical phrasing in expressive performance groups the notes in a piece so as to present a coherent interpretation of its ideas. A performer's tasks include the determining of some viable groupings of the piece that make musical sense, and the communication of this grouping in performance through the manipulation of expressive parameters such as tempo and loudness. The problem we are concerned with is the automatic extraction of phrases – the groupings of notes in a piece.

Phrase structure analysis can begin with the score, or from a performance of a piece. The first focuses on features such as motives, melodies and chord progressions. The latter begins with an expressive performance of a piece. The interpretation is manifested as a set of grouping strategies inherent in the performance. A goal of this paper is to propose a computational approach to automatically extract phrase boundaries from expressive performances based on tempo variations. Our approach finds the best fit sequence of phrase tempo curves using dynamic programming (DP). This paper also explores the relation between tempo variation and phrase structure.

The phrase extraction steps consist of: tempo extraction from audio recordings, tempo smoothing to obtain trajectories, and determination of phrase boundaries from tempo information. We present a Java program for extracting tempo from manually tapping to beat-level onsets. We introduce a DP algorithm for determining the phrase boundaries by curve fitting. We test two quadratic curve types for modeling phrase-level tempo variations: asymmetric concave curves, and splines.

2 RELATED WORK

Most researchers focus on three dimensions of expressive musical performance: tempo, dynamics, and articulation. Gabrielsson [2], Kendall & Carterette [3], Todd [7, 8], and Palmer [5] have found that performers tend to indicate phrase boundaries by lengthening note values at these boundaries, and by increasing the time between successive tones. Similarly, Palmer & Hutchins [6] noted that the phrase is a musical unit that is often demarcated by prosodic cues. Large & Palmer [4] used oscillator models and the product of the probabilities that an onset deviates from expected and that it is late as a measure of phrase boundary likelihood. The oscillator models require initial phase and period information. Cheng & Chew [1] used local maxima in the loudness time series to detect phrases. Our present approach uses DP to fit tempo trajectories using a sequence of quadratic curves so as to determine phrase boundaries. No prior information is required in the latter two techniques.

3 SYSTEM DESCRIPTION

The system consists of two parts: tempo extraction and phrase boundary determination. Figure 1 shows the system diagram. We describe each part in this section.

3.1 Tempo Extraction

We develop a Java program (the tapper in Figure 1) to record the time of a user's tapping for generating the

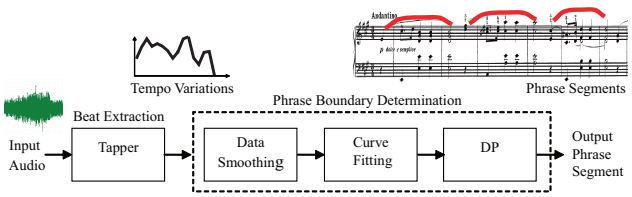

Figure 1. Phrase extraction system diagram

tempo time series from an expressive performance. Figure 2 shows a screenshot of the program. The interface contains three panels, showing a representation of the score with pitch height and onsets (from MIDI input), the amplitude (from the audio signal), and the tempo (calculated from the user's tapping) respectively.

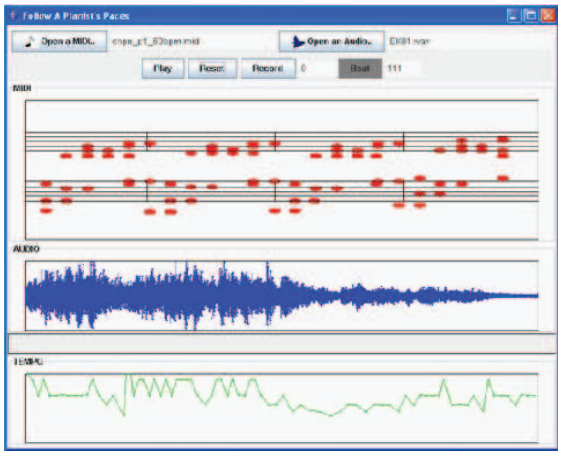

Figure 2. Screenshot of the tapper program

To extract tempo information, one of the authors (Chuan) taps the beat along with the audio recording while reading the score; she taps each performance five times, and we use the average of the five as the beat onset time series. Note that tapping at the beat level performs a first level smoothing of the data. The tempo at each beat is calculated as the inverse of the inter-onset-interval between between the current and the previous beat. The program checks for taps every millisecond.

3.2 Phrase Boundary Determination

This section describes three of our considerations in the phrase boundary determination stage.

3.2.1 Data Smoothing

A consideration in the phrase boundary determination stage is data smoothing. The raw tempo data generated by the process described in Section 3.1 can be noisy. We use a non-causal moving average to smooth the data. Figure 3 shows the tempo data from Evgeny Kissin's performance of Chopin's *Prelude No. 1*, before and after smoothing. We used a window size of 2 bars, i.e., 4 beats in 2/8 time.

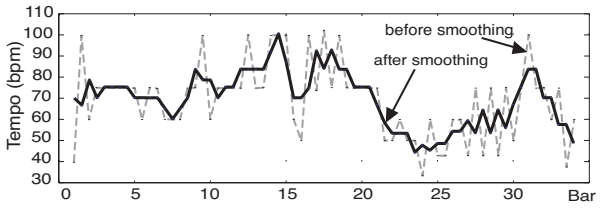

Figure 3. Tempo time series: before and after smoothing

3.2.2 Curve Fitting

We consider two types of curves for modeling the shape of the tempo variation in a phrase. Based on previous findings of tempo slowdowns at phrase boundaries, the curve should consist of two lower ends at the boundaries, with higher values in between. We tested the use of asymmetric concave quadratic curves, and quadratic splines, to model the phrase tempo variations.

The first model, the quadratic curve, possesses four characteristics: (1) it is defined by a degree two polynomial function; (2) the curve is concave, with values in the middle higher than the two ends; (3) the two sides can be asymmetric; and, (4) the peak does not need not to be exactly in the middle. The best-fit curve is determined by the least mean square error (LMSE), a commonly employed measure, between the fitted curve and the original data points. We use quadratic programming to solve the constrained least-squares problem to find the best-fit curve.

Suppose we have the following onset times and tempo values for a phrase, given by $X = \{x_1, x_2, \ldots, x_n\}$ and $Y = \{y_1, y_2, \ldots, y_n\}$ respectively. We use two quadratic functions to model the two asymmetric sides of the curve:

$$\hat{y} = a_1 x^2 + b_1 x + c_1; \qquad \hat{y} = a_2 x^2 + b_2 x + c_2, \quad (1)$$

where $a_1, a_2 \leq 0$ (concavity constraint). The two curves peak and meet at a specific x value, x_p. We iterate through all candidate values, $x_p \in X$, to find the best fit curve. We prune the search space by restricting each equation to descend on only one side of the curve:

$$\frac{b_2}{2a_2} \leq x_p \leq \frac{b_1}{2a_1}. \quad (2)$$

The second model uses the quadratic spline, a piecewise polynomial function with one continuous derivative. A local minimum search determines the boundaries after curve fitting.

3.2.3 Determining Boundaries

We use DP to determine the phrase boundaries from the expressive performance tempo graph. The objective of the DP algorithm is to minimize the sum of LMSE when approximating the tempo time series by a sequence of quadratic curves (or spline curves); in the process, it segments the entire data stream into a sequence of phrases.

The DP algorithm is based on the observation that the optimal objective value (minimum error) for beats 1

through i is the sum of the error of the optimal solutions for beats 1 through j and the curve fit error for beats j through i, for $1 \leq j < i$, as shown in Equation 3:

$$Opt(1,i) = \min_j [Opt(1,j) + Err(j,i)], \qquad (3)$$

where $j = 1, \ldots, i-1$. In implementation, the initial optimal costs for beats 1 and 2 are set to zero, $Opt(1,1) = Opt(1,2) = 0$, because at least three points are needed for defining a quadratic curve. The smallest interval between two boundaries is set at two bars. The DP algorithm is shown in Table 1.

Table 1. DP algorithm for phrase boundary determination

$n = length\_of\_piece$; % in beats
$p = two\_bars$; % minimum phrase size
$Opt(1,a) = 0 \ \forall \ a \in [1,2]$; % initialization
$Pre(b) = 1 \ \forall \ b \in [p,n]$; % initialization
for $i = p+1 : n$
  for $j = 1 : i-p$
    $Opt_j(1,i) = Opt(1,j) + Err(j,i)$;
  end
  $Opt(1,i) = \min_{j \in [1,i-p]} Opt_j(1,i)$;
  $Pre(i) = \arg\min_{j \in [1,i-p]} Opt_j(1,i)$;
end
return $Pre(n), Pre(Pre(n)), \ldots, 1$;

For the spline model, some of the best-fit curves generated by the DP algorithm may be convex due to a lack of shape constraints. In such cases, we search for the local minima to find the phrase boundaries.

4 EMPIRICAL RESULTS AND DISCUSSIONS

We test the algorithms using the performances of Chopin's *Preludes Nos. 1 and 7* by Evgeny Kissin and Arthur Rubinstein – RCA CD recordings – ASIN: B00002DE5F (Kissin) and ASIN: B000031WBN (Rubinstein, 1946).

Figures 4 through 7 show the extracted phrase boundaries for Kissin's and Rubinstein's performances of *Preludes Nos. 1* and *7*, modeled by quadratic curves and splines respectively. We refer to the two settings of the DP algorithms as DP(quadratic) and DP(spline). The vertical dashed lines emanating down from the smooth curves represent the boundaries determined by the system. The vertical dashed lines cutting across the entire plane are the highest level (several layers of phrasings were labeled) phrase boundaries annotated according to the performances by one of the authors, an expert pianist (Chew).

We observe that the algorithms, both DP(quadratic) and DP(spline), retrieve most of the phrase boundaries indicated by the expert. We discover different tempo strategies employed by the two performers on the same piece, for example, slowdowns at the ends of phrases versus slow starts at the beginnings. Some other challenges of determining phrase boundaries from only tempo information are discussed in the following sections.

4.1 Kissin's Chopin Prelude No. 1

Figure 4 shows the results for Kissin's performance of Chopin's *Prelude No. 1*. The DP(quadratic) algorithm successfully finds two, and DP(splines) three, of the four annotated boundaries (bars 8, 16, 24, and 28). In the final phrase, beginning in bar 29, Kissin employs a slow-start strategy in this performance: the start of this final phrase is as slow as the end of the previous phrase. We consider detected boundaries a bar off from the ground truth due to such slow start strategies as correct.

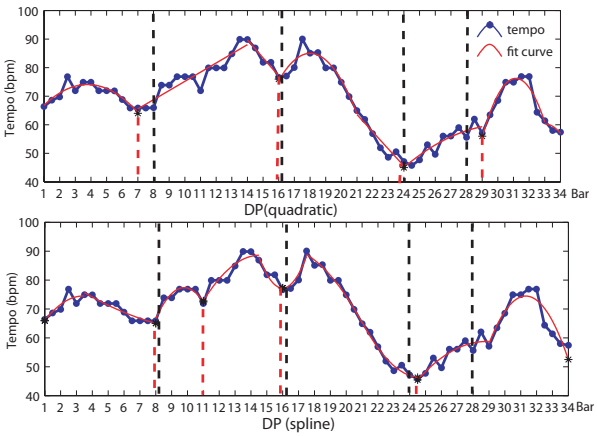

Figure 4. Kissin's Chopin *Prelude 1* boundaries

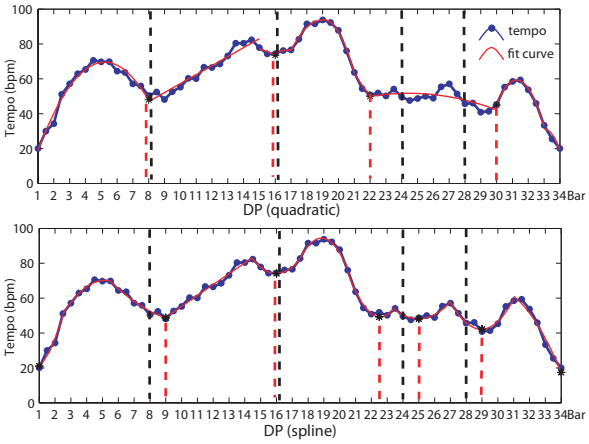

Figure 5. Rubinstein's Chopin *Prelude 1* boundaries

4.2 Rubinstein's Chopin Prelude No. 1

Rubinstein's performance of Chopin's *Prelude No. 1*, shown in Figure 5, exhibits two distinct characteristics in contrast to Kissin's performance. The slowdown at bar 22, before the bar 24 phrase boundary, functions as an agogic accent, an emphasis. This performance also uses multiple

levels of groupings, for example, four two-bar sub-phrases within an eight-bar phrase in bars 1 through 8.

4.3 Kissin's Chopin Prelude No. 7

Figure 6 shows the results for Kissin's performance of Chopin's *Prelude No. 7*. The results, both quadratic and spline, match two of the three annotated boundaries.

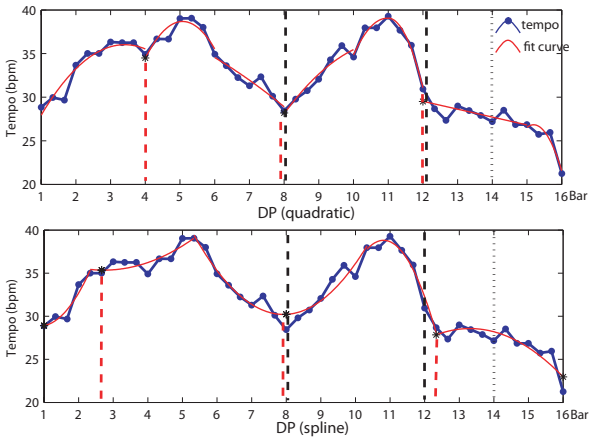

Figure 6. Kissin's Chopin *Prelude 7* boundaries

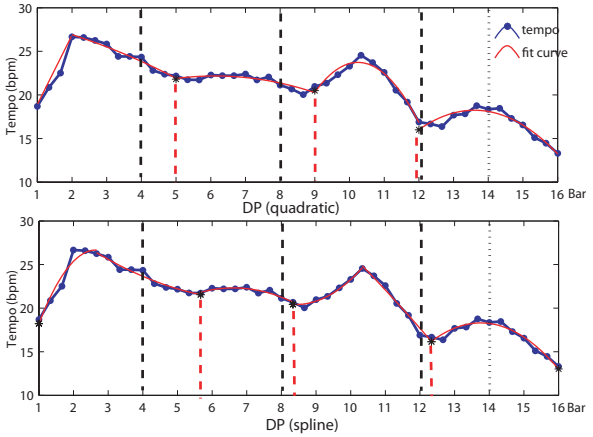

Figure 7. Rubinstein's Chopin *Prelude 7* boundaries

4.4 Rubinstein's Chopin Prelude No. 7

Figure 7 shows the results of Rubinstein's performance of Chopin's *Prelude No. 7*. Comparing Kissin's and Rubinstein's performances, we observe two radically different tempo strategies. In Kissin's performance, two sweeping concave curves are observed in bars (1, 8) and (8, 12) respectively, followed by a relatively level slowdown. In Rubinstein's performance, the tempo shows a predominantly decreasing slope from start to end.

5 CONCLUSIONS AND FUTURE WORK

We have proposed a DP approach for determining phrase boundaries from tempo graphs extracted from expressive performances. The algorithm accurately determined most of the boundaries annotated by an expert in the test data. We discover widely differing tempo strategies. We also uncover some challenges in the determining of phrase boundaries based only on tempo variations: performers sometimes employ multiple levels of grouping strategies, increasing the complexity of phrase boundary analysis; tempo variation alone is sometimes inadequate for determining phrase boundaries (for example, a slow-start strategy can obfuscate the true boundary); and, tempo slowdowns are not always used for segmenting phrases, it is sometimes used for emphasis.

Future work will explore methods for extracting multiple levels of phrase structure and disambiguating slowdown functions in expressive performances. It will also incorporate other features such as dynamics (loudness) in phrase analysis. More manually annotated performances will be tested.

6 ACKNOWLEDGEMENTS

This material is based upon work supported by the National Science Foundation (NSF) under grants No. 0347988 and EEC-9529152. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors, and do not necessarily reflect the views of NSF.

7 REFERENCES

[1] Cheng, E. and Chew, E. "A Local Maximum Phrase Detection Method and the Analysis of Phrasing Strategies in Expressive Performances", *Proc. of Math and Computation in Music*, Berlin, Germany, 2007.

[2] Gabrielsson, A. "Music Performance", *Psychology of Music*, New York: Academic Press, 1999.

[3] Kendall, R. A. and Carterette, E. C. "The communication of musical expression", *Music Perception*, Vol. 8, No. 2, pp. 129-164, 1990.

[4] Large, E. W., and Palmer, C. "Perceiving temporal regularity in music", *Cognitive Science*, 26, pp. 1-37, 2002.

[5] Palmer, C. "Music Performance", *Ann. Rev. of Psychology*, Vol. 48, pp. 115-138, 1997.

[6] Palmer, C. and Hutchins, S. "What is musical prosody?", *Psycology of Learning and Motivation*, Vol.46, Elsevier Press, 2005.

[7] Todd, N. P. M. "The dynamics of dynamics: A model of musical expression", *J. of the Acoustical Soc. of America*, Vol. 91, Issue 6, pp.3540-3550, June 1992.

[8] Todd, N. P. M. "The kinematics of musical expression", *J. of the Acoustical Soc. of America*, Vol.97, Issue 3, pp.1940-1949, March 1995.