A Dynamic Programming Approach to the Extraction of Phrase Boundaries from Tempo Variations in Expressive Performances

USC IMSC
A NATIONAL SCIENCE FOUNDATION ENGINEERING RESEARCH CENTER

Music Computation and Cognition (MuCoaCo) Lab

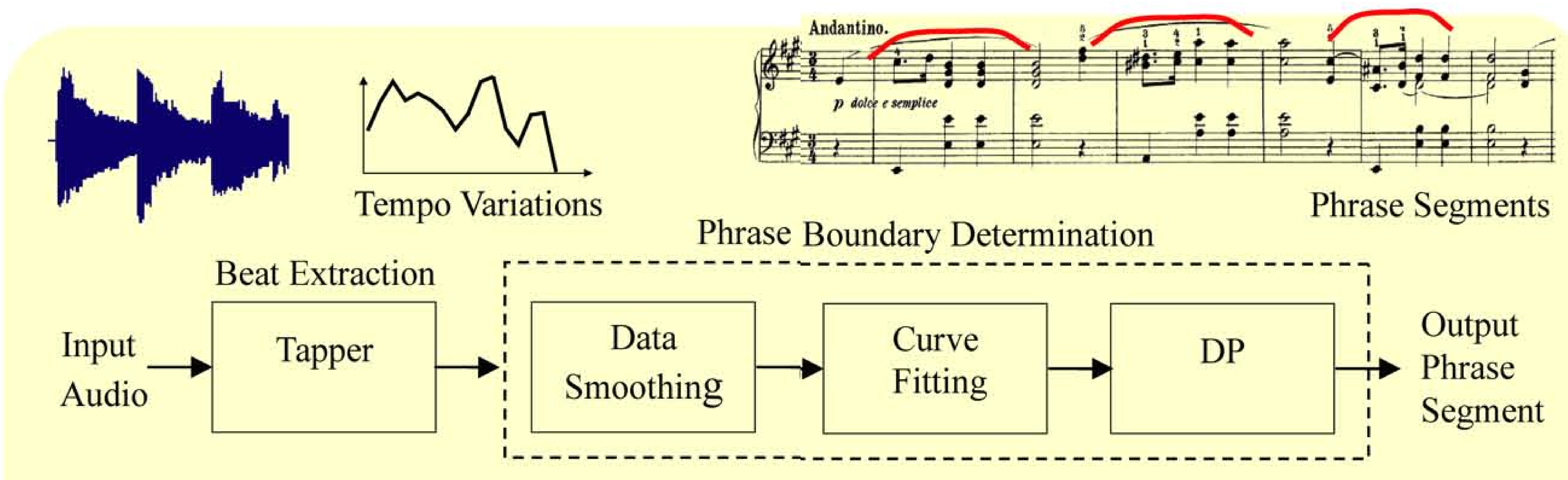
Ching-Hua Chuan
Elaine Chew
{chinghuc, echew}@usc.edu

University of Southern California
Viterbi School of Engineering

Introduction

We present an approach to phrase segmentation that starts with an expressive music performance. Musical phrasing in expressive performance groups the notes in a piece so as to present a coherent interpretation. The problem we are concerned with is the **automatic extraction of phrases – the groupings of notes conveyed in a performance.**

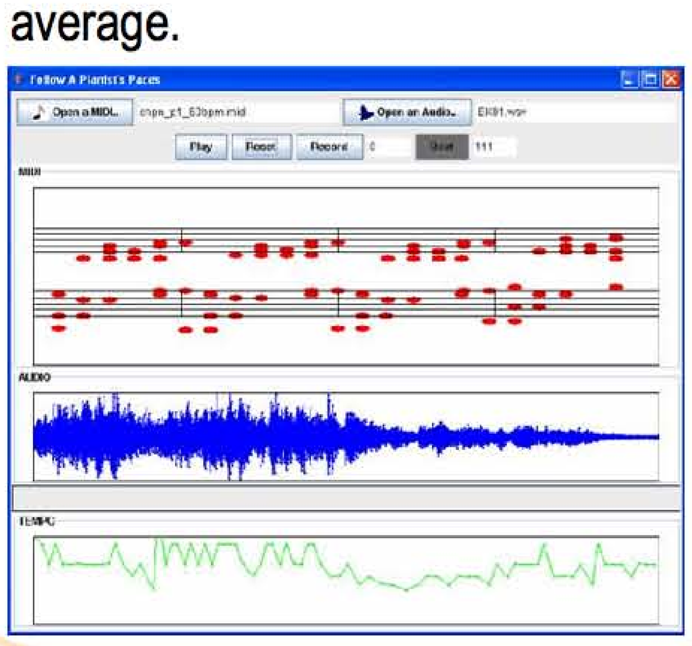
Previous research has shown that phrases are delineated by **tempo speedups and slowdowns**. We propose a dynamic programming algorithm for extracting phrases from tempo information.



Phrase Extraction System Diagram

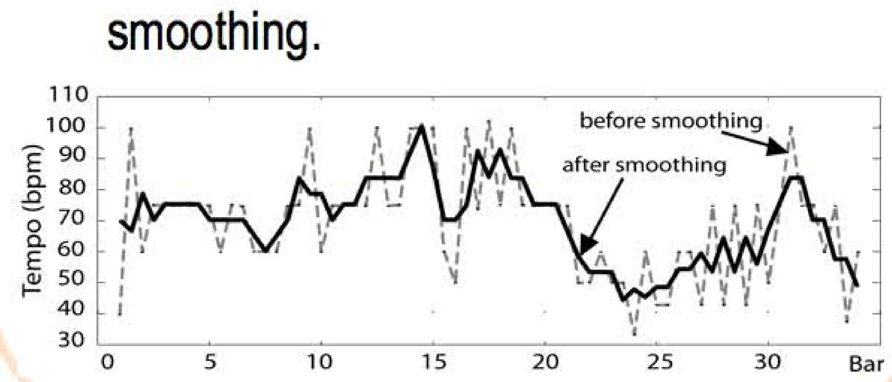
Beat Extraction

We extract tempo information by **tapping the beat** along with the recordings through a Java program. The program checks for taps every millisecond. We tap each performance five times and use the average.



Data Smoothing

The raw tempo at each beat is calculated as the inverse of the inter-onset-interval between beats. A **non-causal moving average** with window size of 2 bars is used for smoothing.



Curve Fitting

We tested the use of **asymmetric concave quadratic curves**, and **quadratic splines**, to model the phrase tempo variations. Each model is referred as DP(quadratic) and DP(spline) respectively.

The best-fit curve is determined by the least mean square error (**LMSE**) between the fitted curve and the original data points.

Related Work

[1] Gabrielsson, A. "Music Performance," *Psychology of Music*, 1999.

[2] Large, E. W., and Palmer, C., "Perceiving temporal regularity in music," *Cognitive Science*, 2002.

[3] Palmer, C., and Hutchins, S. "What is musical prosody?", *Psychology of Learning and Motivation*, 2005.

Dynamic Programming

We use DP to determine the phrase boundaries from the tempo graph. The objective is to minimize the sum of LMSE.

$n = length\_of\_piece$; % 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 $Pre(n), Pre(Pre(n)), \ldots, 1$;

Empirical Results

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

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.

Discussion

We uncover challenges in the determining of phrase boundaries based only on tempo variations:

- performers may employ multiple levels of grouping strategies
- tempo variation alone is sometimes inadequate for determining phrase bdries
- tempo slow-downs are not always used for segmenting phrases

Conclusions / 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.

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.

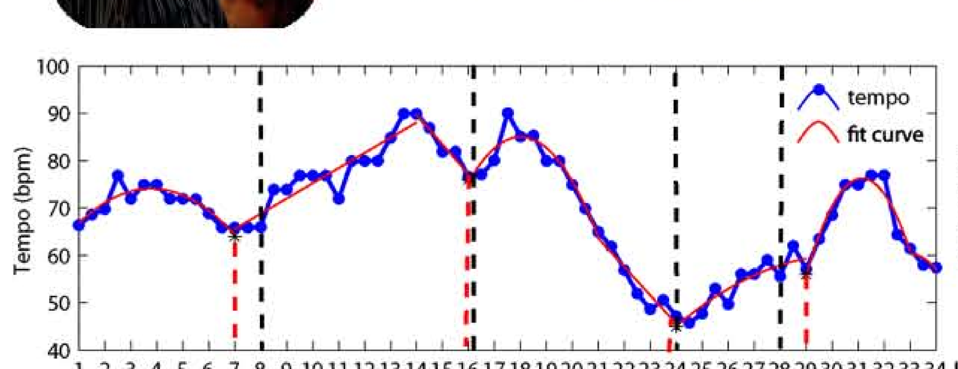
Chopin's *Prelude No. 1*

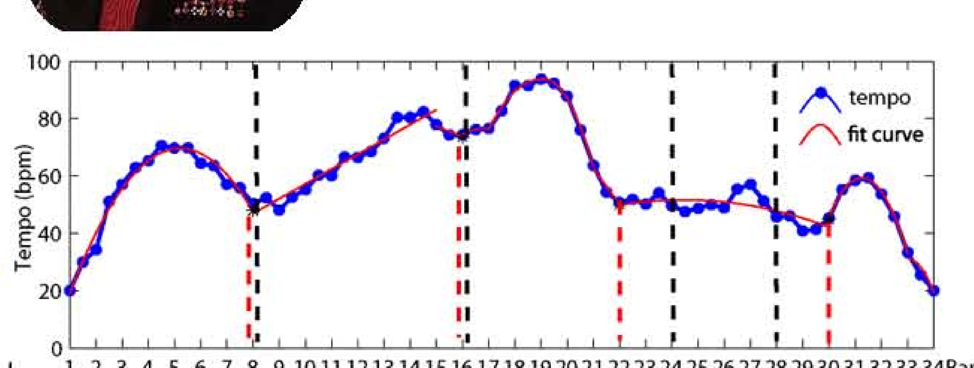
Kissin's Performance

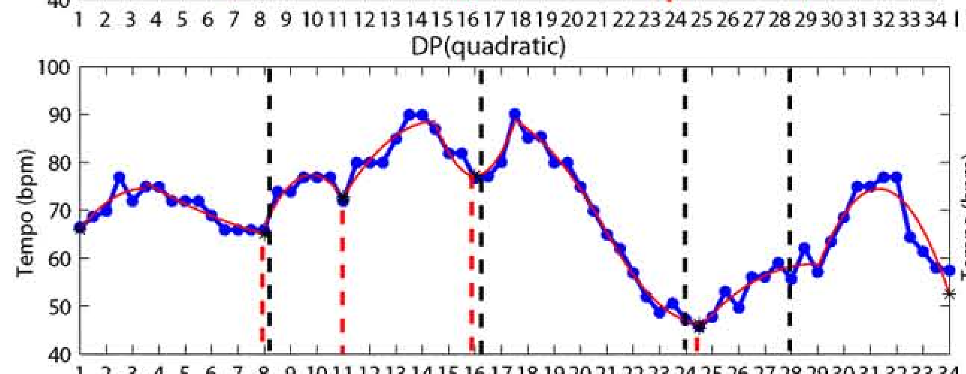
Boundaries at bars 8, 16, 24,and 28.
Slow-start: bar 29.

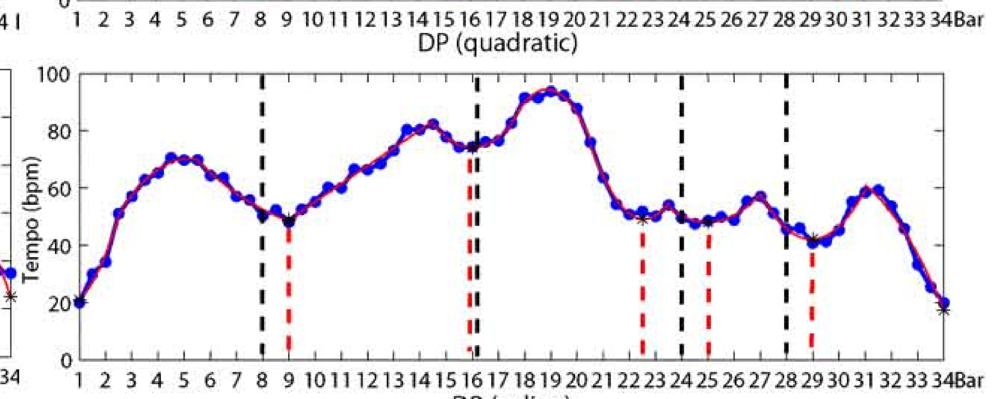
Rubinstein's Performance

Agogic accent: bar 22.
Groupings: four 2-bar sub-phrases.

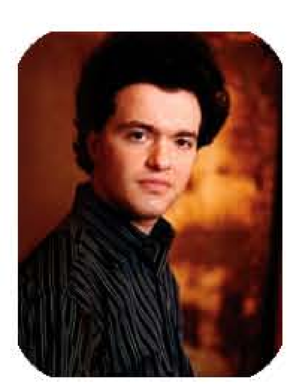








Chopin's *Prelude No. 7*



Kissin's Performance

Boundaries at bars 4, 8, 12, and 14 (weak).

Rubinstein's Performance

Predominant decrease from start to end.