

A Local Maximum Phrase Detection Method for Analyzing Phrasing Strategies in Expressive Performances

Eric Cheng and Elaine Chew

University of Southern California Viterbi School of Engineering

February 24, 2007

Abstract

This paper proposes a Local Maximum Phrase Detection (LMPD) method for the analysis of phrasing strategies in expressive performances. The LMPD method systematically extracts a quantitative representation of phrasing strategy by equating the occurrence of a local maximum in the loudness curve with the occurrence of a phrase or sub-phrase. We further define mathematical descriptors for phrase strength and volatility, and phrase typicality, for comparing phrasing strategies among performances. Phrase strength measures the prominence or clarity of a phrase, and the volatility is defined as the standard deviation of the phrase strengths within a performance. Phrase typicality quantifies the degree to which a phrase peak location is characteristic among the performances polled. We illustrate the LMPD method using preliminary results from its application to eleven commercially available audio recordings of a solo violin Bach Sonata.

1 Introduction

Previous research into expressive phrasing strategies has generally been local in nature. That is, the majority of studies have discussed how performers vary musical parameters within a single phrase or near a single phrase boundary (see, for example, [1], [2], and [4]). While certain local phrasing strategies – such as the clarifying of boundaries with declines in tempo and dynamics – are well documented, relatively little is known about higher level phrasing strategies, i.e., how performers choose to segment a piece into phrases.

In this paper, we aim to develop tools for understanding phrasing strategies from this more global perspective. To that end, we propose the Local Maximum Phrase Detection (LMPD) method, which derives a quantitative representation of phrasing strategies by identifying the number and locations of phrases throughout a performance, and defines mathematical descriptors to quantify the characteristics of each phrase. These methods provide the

means to compare and contrast performance strategies, when combined with pre-annotated score-based phrase segment information.

We extract beat level tempo and loudness data from music audio using manual onset detection and a psycho-acoustic model of loudness. By superimposing author-annotated phrase boundaries over the tempo/loudness data, we find that local maxima in the loudness curve are reliable indicators for phrase or sub-phrase occurrences. This observation leads to the LMPD method, which first equates the occurrence of a local maximum in the loudness curve with that of a phrase or sub-phrase. Then, the strength of definition of each phrase, the *phrase strength*, is calculated as the average loudness increase from the adjoining local minima; and, the *phrase volatility*, is computed as the standard deviation of all phrase strengths. Finally, each phrase is assigned a value, the *phrase typicality*, that quantifies the popularity of its location, among the performances polled.

To demonstrate the efficacy of the LMPD method, we apply the method to performance data obtained from commercially available audio recordings, and present the preliminary results.

2 The Method

This section presents our method for performance analysis, and consists of: (1) a description of our method for tempo and loudness extraction; (2) arguments for focussing on loudness data; and, (3) the proposal of the local maximum phrase detection method, including the extraction of

phrases, and of the mathematical descriptors.

2.1 Data Extraction

We first extract onset times manually using a marking tool in Final Cut Pro. Then, we use the onset times to calculate beat level tempo. To compute loudness data, we first calculate a loudness waveform, in Sonos, using a MATLAB implementation of the PEAQ standard [3]. The waveform is then smoothed using a Gaussian window, and sampled at each onset time, to obtain beat level data. To validate the accuracy of the extracted data, we compare loudness values for a single recording to a manually plotted reference curve representing our own perception. The smoothing window width is optimized so that the loudness data best matches the reference curve.

2.2 The Case for Loudness

We extracted performance data (tempo and loudness) from eleven commercially available audio recordings of the *Andante* movement of Bach's Sonata No. 2 for solo violin. We chose this piece for its regular pulse and unambiguous phrase structure – qualities that simplified both data extraction and analysis. The performances were by Ehnes, Enescu, Grumiaux, Heifetz, Kremer, Menuhin, Milstein (1956, 1975), Mintz, Szeryng, and Szigeti.

To devise a systematic method of phrase detection, we first annotated phrase boundaries using only the score as our guide. Then, as shown in Figures 1 and 2, we superimposed these boundaries over plots showing tempo and loudness data for all

performances to see whether any identifiable patterns arose.

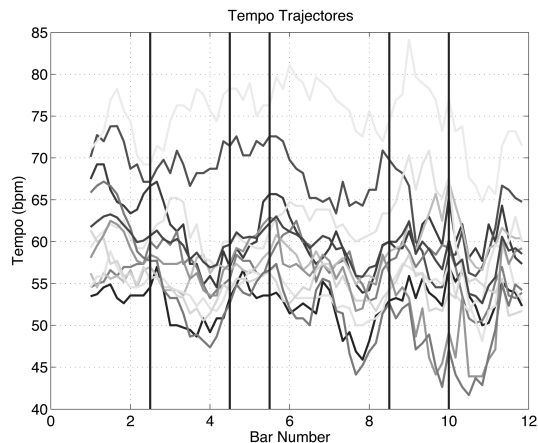


Figure 1: Phrase boundaries superimposed over tempo data.

In the two figures, vertical lines denote starts of phrases, and each trajectory represents a single performer’s performance data.

Observe that the loudness trajectories appear to be more consistently related to the annotated phrase boundaries than their tempo counterparts. In particular, phrases are well characterized by a crescendo/decrescendo arch similar to that mentioned in several past studies, such as [2], [5], and [6]. In some cases, phrases are characterized by two or more sub-arches, suggesting that those performers chose to divide the annotated phrases into sub-phrases.

In contrast, the tempo strategies are less systematically related to the phrase boundaries, with a greater diversity of trajectories. These observations, which were confirmed by analyzing the average inter-performer correlations for the tempo and loudness data ($\bar{r}_{tempo} = 0.4862$ and

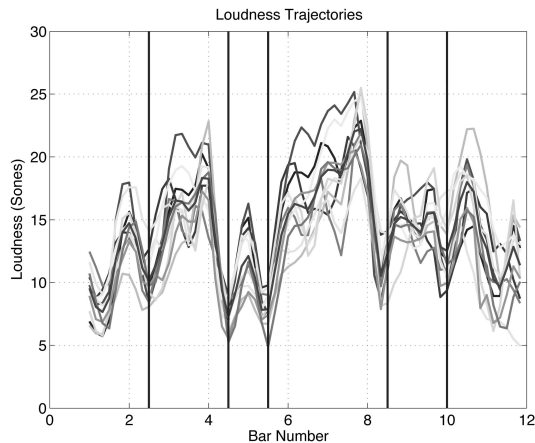


Figure 2: Phrase boundaries superimposed over loudness data.

$\bar{r}_{loudness} = 0.7627$, respectively), led us to conclude that loudness is a more reliable parameter for phrase detection, and in particular, that a crescendo/decrescendo arch is a reliable indicator for the occurrence of a phrase.

2.3 Local Maximum Phrase Detection

If we assume that each phrase is characterized by a crescendo/decrescendo arch, then each phrase should also be associated with a local maximum in loudness. The LMPD method uses this local maximum as a mathematical indicator for the existence of a phrase. The method consists of two steps: (1) record number and locations of local loudness maxima for each performance; and, (2) interpret each local maximum as a phrase or sub-phrase.

The total number of local maxima in a performance provides a global measure of the degree to which a performer highlights

local vs. global phrase structure, while the locations of the local maxima allow us to compare different phrase subdivision strategies. This method also allows us to define additional mathematical descriptors to further quantify the characteristics of phrasing strategy. These are discussed in the next two sections.

2.3.1 Phrase Strength and Volatility

We define the *phrase strength* ($P.S.$) of a phrase to be equal to the average loudness difference between its local maximum and the two adjoining local minima:

$$P.S. = \frac{1}{2} [(M_i - m_j) + (M_i - m_k)] , \quad (1)$$

where M_i is the loudness value of the local maximum and m_j , and m_k are the loudness values of the two adjoining local minima as shown in Figure 3. $P.S.$ values allow us to measure the prominence or clarity of a particular phrase.

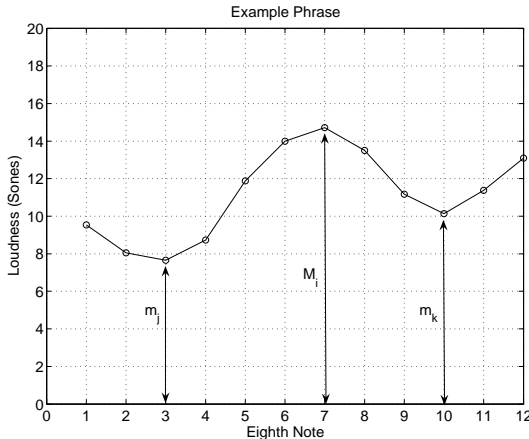


Figure 3: Phrase strength parameters.

The *phrase volatility* ($P.V.$) of a particular performance is defined to be the stan-

dard deviation of all $P.S.$ values in the performance. Thus, the greater the variability in phrase strengths, the greater the phrase volatility.

2.3.2 Phrase Typicality

The *phrase typicality* ($P.T.$) of a phrase quantifies the popularity of its location. It is defined to be the proportion of other performers who also place a local maximum at the location of the phrase in question. Mathematically, the phrase typicality is given by:

$$P.T. = \frac{1}{N - 1} [M(i) - 1] , \quad (2)$$

where $M(i)$ is the total number of performers placing a local maximum at location i , and N is the total number of performers. Thus, the greater the number of performers placing a local maximum at a particular location, the greater the phrase typicality of a phrase at that location. Figure 4 shows how $M(i)$ varies from location to location. In the sample performance trajectory, on the top half of Figure 4, the vertical dotted lines indicate local maxima. The histogram on the lower half of Figure 4 shows the number of performers placing a local maximum at a particular location, equivalent to $M(i)$ in Equation 2. The histogram shows, for example, that the peak at bar 5 is highly typical, and shows up in eleven of the twelve recordings, while a peak at the beat just before it was observed in only one recording.

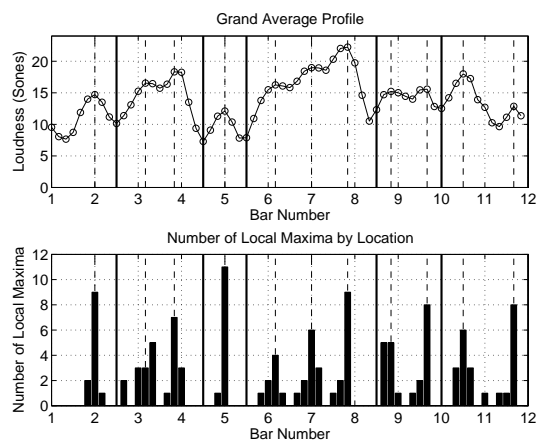


Figure 4: Top: Average performance trajectory. Bottom: Number of performers placing a local maxima at a particular location.

3 Conclusion

In conclusion, we have presented a novel method for the analysis of phrasing strategies in expressive performances by equating the occurrence of a local maximum in loudness with the occurrence of a phrase. Preliminary results suggest that a local maximum in loudness is a meaningful expressive event that reliably indicates the presence of a phrase. We defined mathematical descriptors to quantify the characteristics of each phrase, and comprehensive analyses of the twelve Bach recordings will follow. Until we conduct listening tests, we can only hypothesize about the perceptual significance of these descriptors.

Acknowledgements

This material is based upon work supported by a Frank H. Buck Scholarship, and by the National Science Foundation under grant

No. 0347988. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors, and do not necessarily reflect the views of the National Science Foundation.

References

- [1] E. Cambouropoulos. The local boundary detection model (lbdm) and its application in the study of expressive timing. In *Proc. of the Intl. Computer Music Conf.*, 2001.
- [2] A. Gabriellson. Once again: the theme from mozart’s piano sonata in a major (k.331). In *Action and Perception in Rhythm and Music*, pages 81–103. Royal Swedish Acad. of Music, Stockholm, 1987.
- [3] P. Kabal. *An Examination and Interpretation of ITU-R BS.1387: Perceptual Evaluation of Audio Quality*. McGill University, May 2002.
- [4] J. Langner and W. Goebel. Visualizing expressive performance in tempo-loudness space. *Computer Music J.*, 27(4):69–83, Winter 2003.
- [5] J. Sundberg, A. Friberg, and R. Bresin. Attempts to reproduce a pianist’s expressive timing with director musices performance rules. *J. of New Music Res.*, 32(3):317–325, September 2003.
- [6] N. P. M. Todd. The dynamics of dynamics: a model of musical expression. *J. of the Acoustical Soc. of America*, June 1992.