

A STATISTICAL VIEW ON THE EXPRESSIVE TIMING OF PIANO ROLLED CHORDS

Mutian Fu¹

Guangyu Xia²

Roger Dannenberg²

Larry Wasserman²

¹School of Music, Carnegie Mellon University, USA

²School of Computer Science, Carnegie Mellon University, USA

{mutianf, gxia, rbd, larry}@andrew.cmu.edu

ABSTRACT

Rolled or *arpeggiated* chords are notated chords performed by playing the notes sequentially, usually from lowest to highest in pitch. Arpeggiation is a characteristic of musical expression, or expressive timing, in piano performance. However, very few studies have investigated rolled chord performance. In this paper, we investigate two expressive timing properties of piano rolled chords: *equivalent onset* and *onset span*. Equivalent onset refers to the hidden onset that can functionally replace the onsets of the notes in a chord; onset span refers to the time interval from the first note onset to the last note onset. We ask two research questions. First, what is the equivalent onset of a rolled chord? Second, are the onset spans of different chords interpreted in the same way? The first question is answered by local tempo estimation while the second question is answered by *Analysis of Variance*. Also, we contribute a piano duet dataset for rolled chords analysis and other studies on expressive music performance. The dataset contains three pieces of music, each performed multiple times by different pairs of musicians.

1. INTRODUCTION

Rolled (or *arpeggiated*) chords are notated chords performed by playing the notes sequentially, usually from lowest to highest in pitch. It is a common technique and an integral part of musical expression. Especially, pianists use rolled chords to convey their interpretations of expressive timings. In a very broad sense, every piano chord is rolled since no two notes are played exactly at the same time.

However, very few works have investigated piano rolled chords. As a consequence, when dealing with chords, most *expressive performance* studies stick to the melody or top note, in part due to a lack of theoretical foundations. For example, when analyzing the timing of a chord, researchers usually simply take the onset of a certain note in a chord (e.g., the first note or the highest note) as the onset of a rolled chord [4][13] even though authors realize this is not the best solution. When synthesizing the timing of a chord, people either put the note

onsets of a chord at exactly the same time or decode the onsets of each note individually [6][15]. This situation motivates us to investigate some fundamental properties of rolled chords in order to set a better basis for future expressive performance studies.

We investigate two expressive timing properties of piano rolled chords: *equivalent onset* and *onset span*. Equivalent onset refers to the hidden onset that can functionally replace the onsets of the notes in a chord; onset span refers to the time interval from the first note onset to the last note onset. We compute equivalent onset time and relative location within a rolled chord via local tempo estimation, assuming that local tempo is steady within a few beats. To be more specific, we first estimate a linear mapping (a tempo map) between real performance time and score time for each chord. Then, we compute the intersection between the tempo map and the chord's onset span to compute a hidden equivalent onset. Finally, we compare the equivalent onset with the note onsets of the rolled chord to figure out its relative location. For onset span, we focus on a more fundamental statistical problem: if onset spans are considered random variables, are they drawn from the same distribution, or affected by different chords or performances? We solve this problem by using *Analysis of Variance* (ANOVA). In our case, ANOVA provides a statistical test of whether the means of onset spans of different chords are equal.

The next section presents related work. Section 3 describes a new data set we created for this study. Section 4 presents an important data preprocessing (polyphonic alignment) procedure. In Sections 5 and 6, we show the methodologies for equivalent onset and onset span, respectively. In Section 7, we present experimental results.

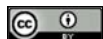
2. RELATED WORK

We review two realms of related work: polyphonic alignment and piano rolled chords. The former is only related to our data preprocessing procedure while the latter is related to the main goal of our study.

2.1. Polyphonic Alignment

Researchers have developed both *online* and *offline* polyphonic alignment algorithms for both audio and symbolic data. Our study uses offline symbolic polyphonic alignment based on the MIDI representation.

For audio-based polyphonic alignment, researchers usually first analyze an audio spectrogram to extract pitch and timing features and then perform an alignment



based on extracted features. Cont [2] uses non-negative matrix factorization for polyphonic pitch analysis and then uses a hierarchical hidden Markov model to achieve the alignment by sequential modeling. Raphael [11] introduces a graphical method to detect latent tempo and current position in score.

Compared to audio-based approaches, symbolic alignment is relatively easy since the target files usually contain accurate pitch and timing information. Bloch and Dannenberg [1] introduce two online algorithms as a part of the first polyphonic computer accompaniment system. Their work uses pitch information and a rating function to find the best fit between performance and score. Hoshishiba et al. [8] propose an offline approach by using dynamic programming and spline interpolation, in which dynamic programming is used to find the maximum match between performance data and score and spline interpolation is used to post-process and improve the result. A more recent research is done by Chen et al. [3], in which two methods are introduced. The first method sorts notes in a MIDI file by their onset and then uses longest common subsequence to map the performance to the score. The second method sets some correctly matched notes as the pivots, separates note sequence by those pivots, and optimizes the result recursively by forward and backward scanning.

2.2. Piano Rolled Chords Study

There are fewer studies related to piano rolled chords. From an analysis perspective, Repp [12] investigates some descriptive properties of arpeggiated chord onsets by using a single piece of music. To be more specific, this study considers the relative onset timing and inter-onset-interval within arpeggiated chords. It compares the results between the performances by students and experts and draws the conclusion that arpeggiating patterns are subject to large individual differences. From the synthesis perspective, Kim et al. [9] predict the onsets of a rolled chord by first estimating the onset of the highest note and then adding intervals for the onsets of succeeding notes.

3. DATASET

Besides investigating the equivalent onset time and onset span of piano rolled chords, we contribute a piano duet dataset for rolled chord analysis and other studies on expressive music performance [15]. The advantage of duet performance is that we are able to access the expressive timing from both parts. The dataset currently contains three pieces of music: *Danny Boy*, *Serenade* (by Schubert), and *Ashokan Farewell* [7]. Each piece contains a monophonic melody part and a polyphonic accompaniment part. For the polyphonic part, the three pieces contain 32, 56, and 245 chords, respectively. Each piece is performed 35 to 42 times by 5 to 6 different pairs of musicians (each pair performed each piece of music 7

times). This dataset is now accessible online via www.cs.cmu.edu/~gxia/data.

4. DATA PREPROCESSING

Before investigating the equivalent onset and onset span of any rolled chord, we have to align the polyphonic piano performance to the score. This task is done in two steps: *forward alignment* and *backward correction*.

Forward alignment: We adopt the online approach used by Bloch and Dannenberg [1] for the forward alignment step. Generally speaking, the algorithm takes a performance as sequential inputs and matches performance notes one-by-one to a reference of sorted chords. At each step of the alignment, it maximizes the number of matched score notes minus the number of skipped score notes.

Backward correction: The forward alignment procedure works well for most music, but may cause a problem when adjacent chords share the same note.

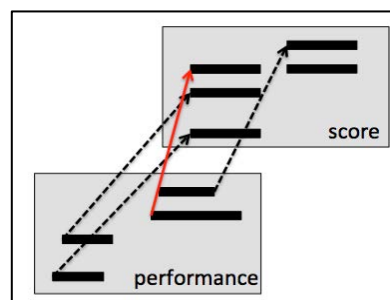


Figure 1. A piano roll illustration of forward alignment procedure.

As shown in Figure 1, dotted arrows represent correct matches while the solid arrow represents the false match. In this case, the top note in the 1st chord is skipped in the performance and the next chord’s 1st performed note happens to share the same pitch with the skipped note. As a consequence, the 1st chord “borrows” the missing note from the 2nd chord. In the worst case, if all the chords share the same note, this mismatch behavior could happen recursively. To address this issue, the backward correction algorithm starts from the last chord and recursively recovers the borrowed notes, if any.

5. EQUIVALENT ONSET

If we replace all the note onsets of a rolled chord by a single onset, where should we place this single onset to let it sound most like the original chord? It is reasonable to assume that this equivalent onset is hidden within the range of the rolled chord’s onset span and has some particular relationship with the onsets. In this section, we first find out the location of the hidden equivalent onset by local tempo estimation. Then we propose two functional approximations to reveal relative onset location within each rolled chord. In the following sections, we

use n to denote the total number of chords of a piece of music and m to denote the total number of performances of a piece of music.

5.1. Absolute Location of Equivalent Onset

If local tempo around rolled chords is stable, equivalent onsets can be linearly interpolated from neighboring onsets. We consider the melody notes within 2 beats of rolled chords and transfer the equivalent onset estimation problem into a beat estimation problem.

Formally, if the current chord index is i , we denote its score onset and equivalent performance onset by $accom_{s_i}$ and \overline{accom}_{p_i} , respectively. We do equivalent onset estimation based on the melody notes whose onsets are within the range of $[accom_{s_i} - 2, accom_{s_i} + 2]$. To be more specific, we first estimate a linear mapping between performance onsets and score onsets of the melody notes within this range. Then, if we denote the slope and the intercept of this linear mapping as α and β , respectively, we can find the equivalent onset by:

$$\overline{accom}_{p_i} = \alpha \cdot accom_{s_i} + \beta \quad (1)$$

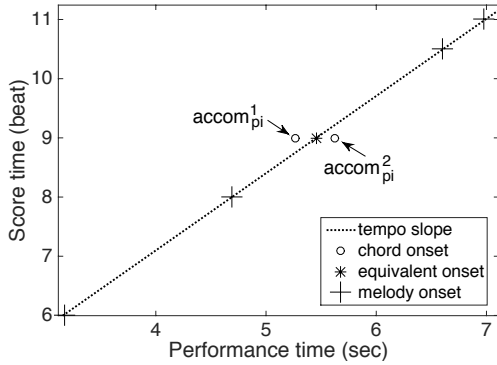


Figure 2. An illustration of equivalent onset estimation by local linear mapping.

This process is illustrated by Figure 2, in which the '+' symbols represent the melody notes and the circle symbols represent accompaniment rolled chord. The line represents the tempo map computed by linear mapping and the star point, on the line at score time 9, represents the equivalent onset computed by equation (1).

5.2. Relative Location of Equivalent Onset

Once the absolute location of equivalent onset is estimated, we present two methods to model its relative location within rolled chords: the ratio model and the constant offset model. For both models, we consider the \overline{accom}_{p_i} computed in the last section as the ground truth and find the models' parameters by minimizing the difference between the models' predictions and the ground truth.

5.2.1 Ratio Model

The ratio model assumes that equivalent onset is decided by the first and last onset of a rolled chord as in the following equation:

$$\overline{accom}_{p_i}'(r) = (1 - r) \cdot accom_{p_i}^1 + r \cdot accom_{p_i}^2 \quad (2)$$

In equation (2), $accom_{p_i}^1$ and $accom_{p_i}^2$ refer to the first and last note onsets in a rolled chord respectively. r is the parameter that characterizes the relative location of equivalent onset. According to the value of r , the equivalent onset can be located as follows:

$r < 0$: equivalent onset is before the first onset of the rolled chord.

$0 \leq r \leq 1$: equivalent onset is between first onset and the last onset of the rolled chord.

$r > 1$: equivalent onset is after the last onset of the rolled chord.

For each piece of music, total number of chords is n and total number of performances is m , we find the optimal r value by equation (3):

$$\hat{r} = \underset{r}{\operatorname{argmin}} \sum_{j=1}^m \sum_{i=1}^n |\overline{accom}_{p_i} - \overline{accom}_{p_i}'(r)| \quad (3)$$

5.2.2 Constant Offset Model

The constant offset model assumes that the equivalent onset is decided by the first onset plus some constant offset s . Formally,

$$\overline{accom}_{p_i}'(s) = accom_{p_i}^1 + s \quad (4)$$

Similar to ratio model, we find the optimal s value by

$$\hat{s} = \underset{s}{\operatorname{argmin}} \sum_{j=1}^m \sum_{i=1}^n |\overline{accom}_{p_i} - \overline{accom}_{p_i}'(s)| \quad (5)$$

6. ONSET SPAN

For onset span, we focus on a more fundamental statistical problem: Do pianists make different interpretations for different chords or performances? As random variables, are all onset spans drawn from the same distribution, or are there different distributions for different chords or performances? In this section, we answer this question by using *Analysis of Variance* (ANOVA). We begin by introducing the basic idea of ANOVA and then link it with our problem step by step.

6.1. One-way ANOVA for Chord Effect

One-way ANOVA can provide a statistical test of whether the means of several groups of data are identical [14]. Formally, if there are n groups indexed by i and μ_i denotes the mean of group i , the null hypothesis and the alternative hypothesis are:

$$H_0: \mu_0 = \mu_1 = \dots = \mu_n \quad (6)$$

$$H_1: \exists i, i': \mu_i \neq \mu_{i'} \quad (7)$$

Generally speaking, one-way ANOVA computes an F-test statistic, which is the ratio of variance between groups to the variance within groups. If different group means are close to each other, this F-test statistics will have a relatively low value and hence retain the null hy-

pothesis. On the other hand, if this F-test statistics is greater than a certain threshold, the null hypothesis will be rejected.

Now let us link this setting to our problem. When checking whether the onset spans of different chords are drawn from the same distribution, each “group” corresponds to a chord and the group members correspond to the onset spans of a particular chord in different performances. In Figure 3, we can see the distributions of the onset span for each chord in *Danny Boy*. The goal is to test whether or not the means of the bars in the boxplot are equal to each other.

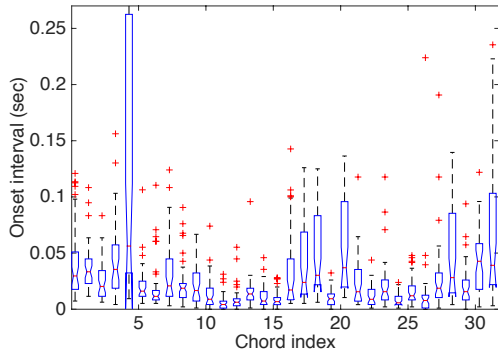


Figure 3. A boxplot of the onset spans of the chords in *Danny Boy*.

Remember each piece of music has n chords and m performances. Therefore, each piece has $N = m \cdot n$ total samples. Referring to the notations in Section 5, the onset span of a rolled chord can be expressed via:

$$t_i = \text{accom}_{p_i}^2 - \text{accom}_{p_i}^1 \quad (8)$$

We use t_{ij} to denote its value in the j^{th} performance. Therefore, the group mean in equation (8) can be computed by

$$\mu_i = \bar{t}_i = \frac{\sum_{j=1}^m t_{ij}}{m} \quad (9)$$

The implementation of one-way ANOVA can be described in the following steps.

First, compute the variation *between* the groups and record its degree of freedom.

$$SS_{btwn} = \sum_{i=1}^n [\bar{t}_i - \bar{t}_{ij}]^2 \quad (10)$$

where $\bar{t}_i = \frac{\sum_{j=1}^m t_{ij}}{m}$, $\bar{t}_{ij} = \frac{\sum_{i=1}^n \sum_{j=1}^m t_{ij}}{N}$. The degree of freedom of SS_{btwn} , $df_{btwn} = n - 1$.

Second, compute the variation *within* individual samples and record its degree of freedom,

$$SS_{within} = \sum_{i=1}^n \sum_{j=1}^m t_{ij}^2 - \sum_{i=1}^n \frac{(\sum_{j=1}^m t_{ij})^2}{m} \quad (11)$$

The degree of freedom of SS_{within} , $df_{within} = N - n$.

Third, compute the F-test statistics by:

$$MS_{btwn} = \frac{SS_{btwn}}{df_{btwn}} \quad (12)$$

$$MS_{within} = \frac{SS_{within}}{df_{within}} \quad (13)$$

$$F = \frac{MS_{btwn}}{MS_{within}} \quad (14)$$

Finally, compare this F-test statistic against a certain threshold to decide whether or not reject the null hypothesis.

6.2. Repeated-measurement One-way ANOVA for Chord Effect

The previous section considered whether different chords have different onset spans. However, an important assumption when using one-way ANOVA is that samples from different groups are independent. In our case, each piece of music is performed by 5 or 6 different pairs of students. Chords played by the same person are clearly correlated. To eliminate the dependent factors produced by same performers, we use repeated-measurement ANOVA to adjust our results.

The general logic of repeated-measurements ANOVA is similar to independent one-way ANOVA. The difference between those two methods is that repeated-measurements ANOVA removes variability due to the individual differences from the within group variance. This process can be understood as removing between-sample variability, and only keeping the variability of how the sample reacts to different conditions (chords). We point readers to Ellen and Girden’s book [5] for more detailed descriptions.

6.3. ANOVA for Performance Effect

Section 6.1 and 6.2 presented the method to inspect whether pianists make different interpretations on onset span for different chords. Following a very similar procedure, if we just exchange the index of i and j in 6.1 and keep everything else the same, we can inspect whether onset spans are interpreted differently for different performances.

7. EXPERIMENTAL RESULTS

7.1. Equivalent Onset

7.1.1 Ratio Model

Figure 4 shows the results of the ratio model. In the figure, the x -axis represents the ratio parameter r and the y -axis represents the relative difference (residual) between model estimated equivalent onset and the ground truth computed via local tempo estimation. Therefore, small numbers indicate better results. Each line corresponds to a piece of music. We see that the optimal r values are all within the range of 0 to 1, indicating that the equivalent onset consistently lies within the range of note on-

sets. The optimal values are 0.42 for *Danny Boy*, 0.13 for *Ashokan Farewell*, and 0.78 for *Serenade*.

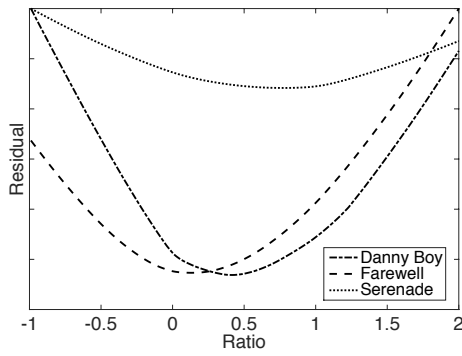


Figure 4. Result of the ratio model.

7.1.2 Constant Offset Model

Similar to Figure 4, Figure 5 shows the results of the constant offset model. The only difference is that the x -axis now represents the constant offset parameter s . We see that the optimal s values are all within the range from 0 to 20 milliseconds. The optimal values are 16 milliseconds for *Danny Boy*, 1 millisecond for *Ashokan Farewell*, and 17 milliseconds for *Serenade*. Compared to the ratio model, the optimal value for constant offset model is more consistent.

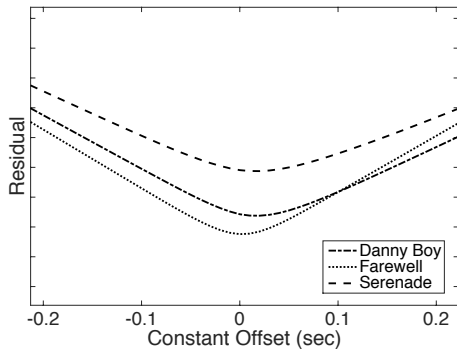
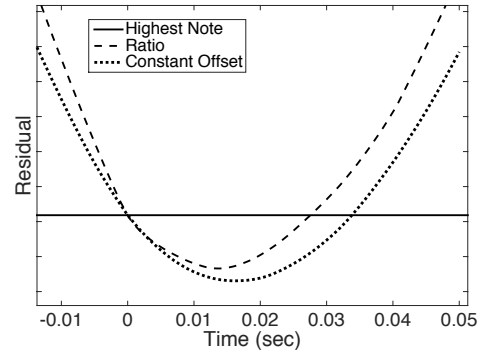


Figure 5. Result of the constant onset model.

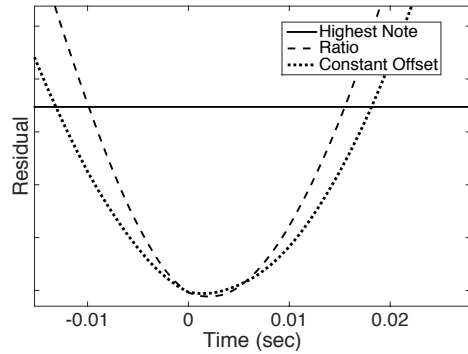
7.1.3 Comparison with Highest Note Model

In most expressive performance studies, people use the highest note onset as the equivalent onset, which we refer to as the “highest note model.” In this section, we compare the results of the ratio model and constant offset model with the highest note model.

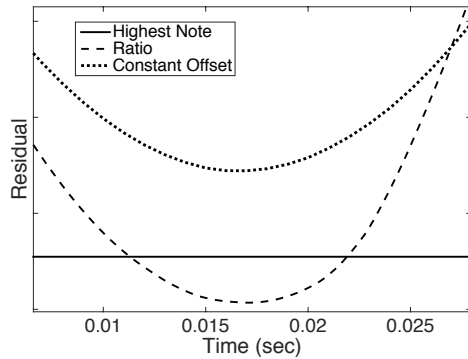
Figure 6 shows this comparison between different models, in which each sub-graph represents a piece of music. Again, smaller number means better prediction. Here, we also map the x -axis value of the ratio model to seconds by multiplying the ratios by the average onset spans. We see that for all the pieces, the ratio model gives better predictions than the highest note model. The constant offset model also does a good job on *Danny Boy* and *Ashokan Farewell* but does not outperform the highest note model for *Serenade*.



(a) Model comparison: *Danny Boy*.



(b) Model comparison: *Ashokan Farewell*.



(c) Model comparison: *Serenade*.

Figure 6. Model comparison of three songs.

7.2. Onset Span

For onset span experiments, we just show the one-way ANOVA table since the repeated-measurement adjustments call for extra notations but give us the same conclusions. Table 1 shows the result of the one-way ANOVA on different chords of *Danny Boy*. Similar to the result of *Danny Boy*, *Ashokan Farewell* and *Serenade* all have the F-test statistics much larger than the thresholds. This indicates that differences between group means are significant. Therefore, we see that not all chords are drawn from the same distribution. In other words, musicians make different interpretations for onset spans of different chords.

Variable	<i>SS</i>	<i>df</i>	<i>F</i>	<i>p</i>
Between	1.6762	31	7.98	4.29×10^{-32}
Within	8.8879	1312		

Table 1. ANOVA for chord effect.

Table 2 shows the result of the one-way ANOVA on different performances of *Danny Boy*. Again, we get similar results for *Ashokan Farewell* and *Serenade*, which all have a F-test statistic not big enough to reject the null hypothesis. This indicates that the differences between group means are *not* significant. Therefore, we see that the interpretations for the same chord’s onset span across different performances are relatively consistent.

Variable	<i>SS</i>	<i>df</i>	<i>F</i>	<i>p</i>
Between	0.2752	41	0.85	0.7383
Within	10.289	1302		

Table 2. ANOVA for performance effect.

8. CONCLUSION AND FUTURE WORK

In conclusion, we create a database to investigate two expressive timing properties of rolled chords in order to set a theoretical basis for future expressive performance studies. We examined three models to characterize the relative location of equivalent onset within rolled chords. The ratio model outperforms the other models for all pieces of music including the highest pitch model used in most research. We also studied onset span. We see that differences are not merely random; musicians use different interpretations for different chords and the interpretation for the same chord across different performances are relatively consistent.

This suggests that in future expressive performance studies, in order to synthesize a rolled chord properly, we can use the equivalent onset as the anchor point (instead of the onset of the highest pitch) and consider the onset span as an important parameter. Although our ratio model improves upon the highest pitch model, the best ratio is different for different pieces and the absolute location of equivalent onset is still based on estimation. This suggests that in future work we should either look for a way to predict the ratio for a given piece of music, or more likely, that we should look for an even better model by combining objective and subjective evaluations.

9. REFERENCES

[1] J. Bloch and R. Dannenberg, “Real-time Computer Accompaniment of Keyboard Performances,” *Proceedings of the International Computer Music Conference*, pp. 279-290, 1985.

[2] A. Cont, “Realtime Audio to Score Alignment for Polyphonic Music Instruments, Using Sparse Non-negative Constraints and Hierarchical HMMs,” *Proceedings of the International Conference on*

Acoustics, Speech and Signal Processing, pp. 245-248, 2006.

[3] C. Chen, J. Jang and W. Liou, “Improved Score-Performance Alignment Algorithms on Polyphonic Music,” *Proceedings of the International Conference on Acoustics, Speech and Signal Processing*, pp. 1365-1369, 2014.

[4] P. Desain and H. Honing, “Does Expressive Timing in Music Performance Scale Proportionally with Tempo?” *Psychological Research*, pp. 285-292, 1994.

[5] R. Ellen and E. Girden, *ANOVA: Repeated Measures*, Sage Publications, 1992.

[6] S. Flossmann, M. Grachten and G. Widmer, “Expressive Performance Rendering With Probabilistic Models,” *Guide to Computing for Expressive Music Performance*, pp. 75-98, 2012.

[7] J. Galway and P. Coulter, *Lengends*, Hal Leonard, 1997.

[8] T. Hoshishiba, S. Horiguchi and I. Fujinaga, “Study of Expression and Individuality in Music Performance Using Normative Data Derived from MIDI Recordings of Piano Music,” *Proceedings of the International Conference on Music Perception and Cognition*, pp. 465-470, 1996.

[9] T. Kim, F. Satoru, N. Takuya, and S. Shigeki, “Polyhymnia: An Automatic Piano Performance System With Statistical Modeling of Polyphonic Expression and Musical Symbol Interpretation,” *Proceedings of the International Conference on New Interfaces for Musical Expression*, pp. 96-99, 2011.

[10] A. Kirke and E. Miranda, *Guide to Computing for Expressive Music Performance*, Springer Science & Business Media, 2012.

[11] C. Raphael, “A Hybrid Graphical Model for Aligning Polyphonic Audio with Musical Scores,” *Proceedings of the International Conference on Music Information Retrieval*, pp. 387-394, 2004.

[12] B. Repp, “Some Observations on Pianists’ Timing of Arpeggiated Chords,” *Psychology of Music*, pp. 133-148, 1997.

[13] B. Repp, “Relational Invariance of Expressive Microstructure across Global Tempo Changes in Music Performance: An Exploratory Study,” *Psychological Research*, pp. 269-284, 1994.

[14] B. Tabachnick and L. Fidell, *Using Multivariate Statistics*, Haper and Row, 2001.

[15] G. Xia and R. Dannenberg, “Duet Interaction: Learning Musicianship for Automatic Accompaniment,” *Proceedings of the International Conference on New Interface for Musical Expression*, 2015.