COMPUTATIONAL ANALYSIS AND MODELING OF EXPRESSIVE TIMING IN CHOPIN MAZURKAS

Zhengshan Shi
Center for Computer Research in Music and Acoustics, Stanford University
kittyshi@ccrma.stanford.edu

ABSTRACT

Performers' distortion of notated rhythms in a musical score is a significant factor in the production of convincingly expressive music interpretations. Sometimes exaggerated, and sometimes subtle, these distortions are driven by a variety of factors, including *schematic* features (both structural such as phrase boundaries and surface events such as recurrent rhythmic patterns), as well as relatively rare *veridical* events that characterize the individuality and uniqueness of a particular piece. Performers tend to adopt similar pervasive approaches to interpreting schemas, resulting in common performance practices, while often formulating less common approaches to the interpretation of veridical events. Furthermore, some performers choose anomalous interpretations of schemas. We present a machine learning model of expressive performance of Chopin Mazurkas and a critical analysis of the output based upon statistical analyses of the musical scores and of recorded performances. We compare the timings of recorded human performances of selected Mazurkas by Frédéric Chopin with performances of the same works generated by a neural network trained with recorded human performances of the entire corpus. This paper demonstrates that while machine learning succeeds, to some degree, in expressive interpretation of schemata, convincingly capturing performance characteristics remains very much a work in progress.

1. INTRODUCTION

Performers of classical music typically interpret a score's symbolic music notation as a basis of performance. This interpretive transformation from symbols to musical sound demands decisions regarding inherently imprecise or vague symbols such as dynamic and tempo markings. Furthermore, performers often divert from strict interpretations of precise symbols such as quantized rhythms in order to provide a sense of musical shape and direction. Expressive timing is a particularly important aspect of performance, with temporal deviations of tempi and distorted rhythms [1] to indicate structural demarcations, express implied affective [2] and articulate stylistic conventions. These interpretive performance decisions are often made with little conscious thought reflecting internalized notions of traditional performance practices and schemas.

The complexity and multidimensionality complicit in the creation of an expressive musical performance has made the task a rich domain for theoretical analysis and computational modeling. Prior studies include analysis-derived rule-based methods such as the KTH model [3], as well as machine learning approaches dating back to Widmer's inference of note-level performance principles based on Sonatas by Mozart [4]. Statistically derived rules include historically rooted schematic tendencies such as the note inégales, arching tempo curves, and cadential ritard were encapsulated in the KTH model. Some generalized schematic rules, such as the tendency to perform a note staccato if the note is repeated immediately, were observed both in KTH and in Widmer's machine learning model.

More recent novel data-driven approaches including both linear [5] and nonlinear [6, 7] methods have been developed to model expressive performance by extracting basis functions (i.e. features) of each note. These features include note, metrical position, dynamic, and tempo markings. Recent efforts apply hierarchical attention networks [8] and conditional variational RNNs [9] to generate expressive piano music performances.

Our goal here is to examine computational models of expressive timing. As noted, performers rarely play metronomically but rather introduce more or less subtle nuances to vary performed durations. For example, most performers tend to slow their tempo in response to major structural breaks [10]. Repp [1] studies patterns of expressive timing over 115 performances of a same piece and suggests independent timing strategies that can describe each pianist's timing pattern. Chew [11] reveals extreme pulse elasticity as musical *tipping points*. Peperkamp et al. [12] propose ways to formally represent relative local tempo variations in a vector space.

We aim to understand how a neural-network-based system generalizes performance practices and compositional style given multiple performances of each of the Mazurkas in our corpus. We train a neural network to predict the tempo curve of each Mazurka. We then analyze expressive timing by comparing human performed Mazurkas to computer generated performances. We observe that while machine learning generalizes key schematic performance practices, it is less successful in capturing veridical performance characteristics.

2. SCHEMATA AND STATISTICAL PERFORMANCE ANALYSIS

Schemata are prototypical melodic, harmonic, or rhythmic/metric characteristics that constitute defining attributes of a particular style or genre [13]. Some schemata, such as the harmonic progression at the approaches to cadential phrase endings, have evolved as pervasive attributes of a common musical language. Chopin's Mazurkas share schemata. They are all composed in triple meter, with regular phrase lengths typically comprised of short motivic units of one or two measures. Along with signature stylistic attributes they also share particular performance practice traditions. In this section we focus on the evolution of schematic features in performance.

2.1 Data

CHARM's Mazurka Project [1] comprises a collection of approximately 3000 individual recorded performances of Mazurkas composed by Frédéric Chopin. Kosta et al. [14] augmented the recordings from the project with automatically aligned score-beat positions, loudness values, as well as positions of expressive markings. The resulting dataset, named as "MazurkaBL", contains 44 Mazurkas with 2000 performances. Sapp [15] has manually annotated 5 Mazurkas with around 300 performances. For each performance, beat times were recorded. These annotations, as well as "MazurkaBL" were used as data for the study.

2.2 Statistical Analysis of Rhythmic Schemata in the Mazurkas

Particular rhythmic patterns characterize the Mazurka, as evident in the frequency of pattern occurrences across the corpus. The ten most recurrent rhythmic patterns are summarized and illustrated in Figure 1.

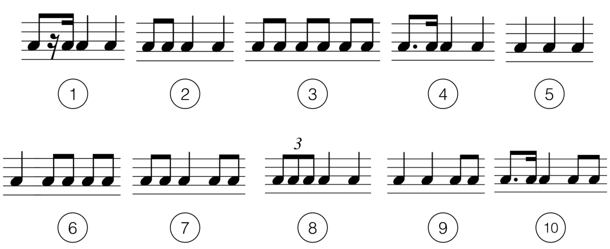


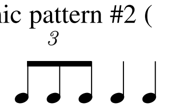
Figure 1: The ten most frequently occurring rhythmic patterns in Chopin's Mazurkas in order of prominence.

We define the *Mazurka Quality* as a beat that is accentuated by temporal elongation repeatedly across most instances of a given Mazurka. Mazurka performances are often characterized as having a "stretched" second or third beat of the measure, at the durational expense of the downbeat [16]. We observed this *short-long* pattern in some mazurkas. However, we also observed that the elongation of the Mazurka Quality was not always compensated for by shortening the downbeat. To validate this, we observed how each pianist executes a pattern on each measure, comparing the duration of the downbeat to the other beats in that measure. Alas, across all performances in the dataset, we observed that only 47.93% of the downbeats were shortened (as compared with the second beat).

[1] http://www.mazurka.org.uk

Pattern #	% of 1st beat shortened
1	45.30%
2	**67.74%**
3	37.11%
4	50.29%
5	56.28%
6	22.00%
7	58.08%
8	**82.08%**
9	39.23%
10	33.36%

Table 1: Percentage of the first beat shortened over all recorded performances of our corpus for the 10 most recurring Mazurka rhythmic patterns. Note that in most performances, the first beat is shortened in rhythmic patterns #2 and #8, indicating a schemata interpretive performance practice.

This suggests that the *short-long* pattern appears only in specific rhythmic patterns. We then examined the tempo curve where human pianists played the above 10 rhythmic patterns respectively. Table 1 summarizes a comparison of how the duration of the first beat of each rhythm is altered in the recorded performances compared with the second beat. These 10 rhythmic patterns comprise over 70% of all rhythmic patterns in 44 Mazurkas in MazurkaBL. We see that in column 2 of Table 1, for rhythmic pattern #2 () and rhythmic pattern #8 (), there is a great percentage of downbeats being shortened (67.74% and 82.08%).

2.3 High Correlation Sections

We examined musical phrases where performers have the highest agreement. We calculated the Pearson's correlation coefficient (PCC), a statistic that measures linear correlation between two variables (or two sets of numbers). This method was previously used by Sapp [15] to represent similarities between performers of Mazurkas. Pearson's correlation coefficient is defined as:

$$\frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^{n}(x_i - \bar{x})^2 \sum_{i=1}^{n}(y_i - \bar{y})^2}} \tag{1}$$

The equation means to divide the covariance of the two variables by the product of their standard deviations. This results in a number ranging from -1 to 1 where -1 indicates negative correlation, 0 indicates no linear correlation, and 1 indicates positive correlation.

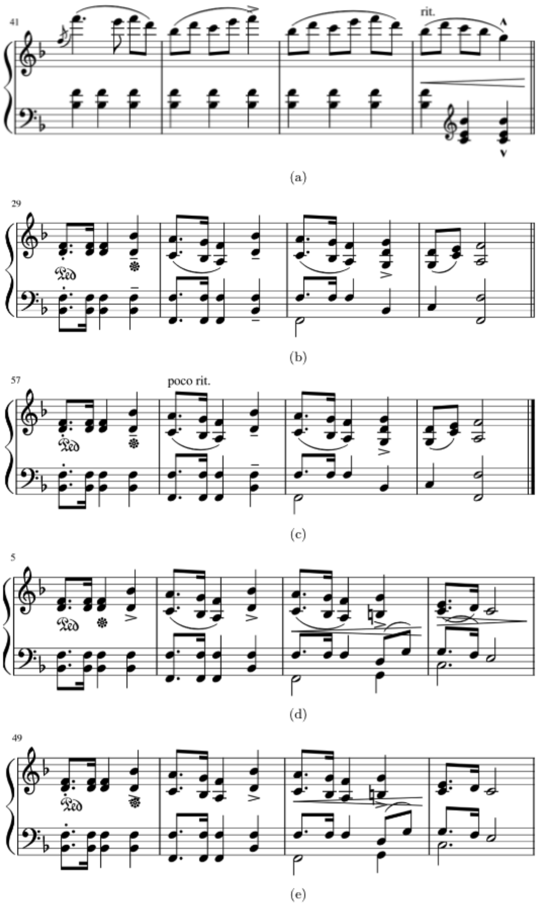


Figure 2: Phrases with the highest correlation in tempo (in order) in Op.68 No.3 performed by 53 pianists.

We looked at Op.68 No.3. As in Figure 2, we found that phrase ends have the highest correlation. Figure 2(a) is the end of C section, 2(b) is the end of the A section, 2(c) is the very end, 2(d) is the end of phrase 1, and 2(e) is the end of phrase 1 after the C section. The Pearson's correlation coefficient of phrase (a) through (e) are: 0.84, 0.84, 0.79, 0.76 and 0.72. When we plotted the tempo of how 53 different pianists playing the highest correlation phrase in Figure 3, we found that all have similar trends: the tempo of the phrase reached climax at the beginning, and then gradually slowed down towards the end. A similar situation was found in Op.24 No.2. The tempo curves of the phrase reached highest at the beginning, then dropped tremendously at the second half of the phrase.

3. VERIDICAL EVENTS IN MAZURKA PERFORMANCES

We borrow Bharucha's [17] distinction between *schematic* events and *veridical* events. A *veridical* event is a musical occurrence characterized by something unexpected within the context of the work. This salient characteristic—whether rhythmic, melodic, harmonic, textural, articulatory or a combination thereof—is typically relatively unique and rare in the specific work, and often is noticeable and attention grabbing. As opposed to a schema, veridical events are less likely to have broadly shared prevalent performance practices.

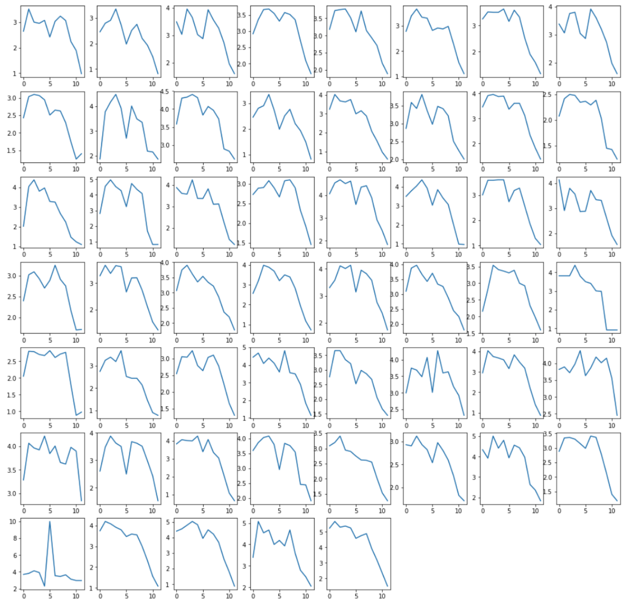


Figure 3: Tempo curves of 53 pianists playing the phrase with the highest correlation in Op.68 No.3. The Y axis represents beat-per-second.

3.1 Unexpected Change in Harmony

Unexpected changes in harmony often cause veridical events. In the opening passage of Op.24 No.1 (Figure 4), the signature rhythm () appears for 6 times (beats 1–3, 7–9, 13–15, 19–21, 31–33, and 37–39). They are rhythmic pattern #1. According to the schema in section 2, they should be performed as a lengthening of the first beat and a shortening of the second or third beat. But performers did not all follow this schema. For the repeating motif in beats 1–3, 13–15, as well as 37–39, most pianists performed the first beat long and the second beat short. However, in beats 7–9, 19–21, 31–33, most pianists changed their Mazurka Quality to lengthen the third beat and shorten the first beat. This is due to a change of harmony in these beats. For example, in beat 9 and beat 33 the piece goes to a vii-th chord that makes the F# in the top voice lead to the G on the next measure. In beat 21, there is an accidental of C# that leads to D. These leading actions cause the elongation of the third beat, rather than the downbeat.

Another example is in Op.63 No.3. Figure 5 and Figure 6 show two phrases that appear at the end of the A section, and at the end of the A' section of Op.63 No.3. According to the trends we summarized in section 2, the phrases that are located at the end of a section are usually the highest correlated phrases among all pianists, as the tempo curves are usually gradually going down. However, these two phrases are the least correlated phrases among different pianists in the whole piece. When we performed a harmonic analysis, we found that there is a secondary dominant chord in the middle of both phrases. The secondary

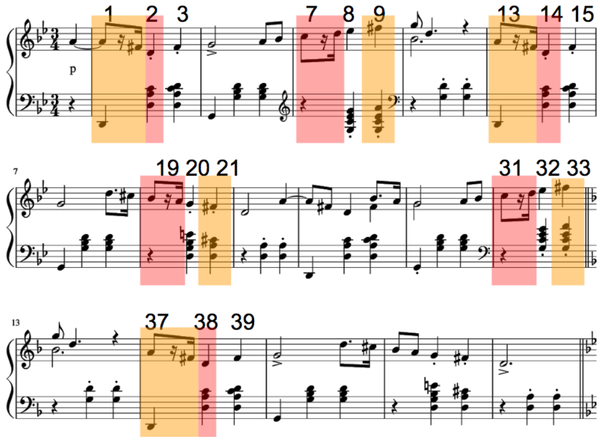


Figure 4: The first theme of Op.24 No.1. Veridical events happen on highlighted notes where there are unexpected changes in harmony. Yellow rectangles mark a lengthening of the beat and red rectangles mark the shortening of the beat.

Figure 5: The low correlation phrase in Op.63 No.3. There is a secondary dominant chord in measure 7.

dominant chords change the harmonic color. In addition, the last chord of each phrase is a common tone diminished chord. Thus the phrase becomes a veridical event.

3.2 Rubato

In Op.24 No.2, the least correlated phrase among 68 pianists is the one with a “rubato" mark on it, as in Figure 7. This phrase appears after the B phrase, serving as a transitional phrase. When Chopin marks rubato, it typically departs from the Mazurka schema and is thus a veridical event, as there is no strict rule about how to perform this excerpt. Pianists usually play the phrase with their own interpretations.

4. COMPUTATIONAL MODEL OF EXPRESSIVE PERFORMANCE

Can machine play music as expressively as humans do? If so, to what extent? This section describes a computational model to synthesize expressive piano music performances. The motivation is mainly to model the complex

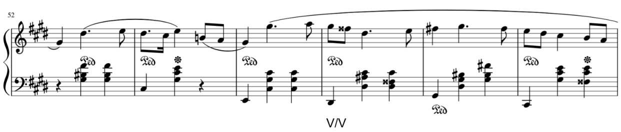


Figure 6: The low correlation phrase 2 in Op.63 No.3. There is a secondary dominant chord in measure 55.

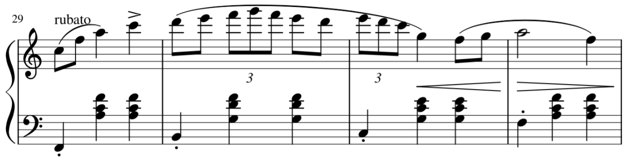


Figure 7: The least correlated phrase in Op.24 No.2.

relationship between score properties and tempo in the performance. The goal is to understand how a neural-network-based system generalizes performance practices and compositional style given multiple performances of each of the Mazurkas in the corpus.

4.1 Input Features

We used MusicXML encoding of the Mazurka scores. Most computational systems of expressive performances take a sequence of note features extracted from MusicXML as input. However, for practical reasons MusicXML format does not readily identify simultaneities. For example, a chord is represented as a sequence of note tokens. Since our goal is to study expressive timing in Mazurkas, especially on how beats are grouped and emphasized, it is important to capture such metrical relationships in the encoding. As a result, we used beat-based features (i.e., features for each beat, rather than for each note).

We first extracted note information on a MusicXML file using *partitura* [18]. For each metrical position, we extracted the following features: highest and lowest notes within the beat, number of simultaneous notes within the beat, the rhythmic pattern of the beat (i.e., triplet, two eighth notes, one quarter note, etc.), articulation (accent and staccato) markings in the beat, metrical phase (i.e., first, second, or third beat in the measure), indicator of the start beat of a phrase, indicator of the final beat of a phrase. The maximum and minimum pitch are represented numerically between 0 and 1, while the rest of the features are represented by one-hot vectors.

4.2 Output Features

Piano allows for expressive variation in timing, dynamics, and articulation [1]. The output features are velocity, tempo, and articulation.

The velocity is a numeric value between 0 and 1, corresponding to the sone values in the dataset. The beat tempo is first calculated as the reciprocal of the beat interval such that

$$\text{tempo}_i = \frac{\text{IBI}_i^{\text{score}}}{\text{IBI}_i^{\text{perf}}} = \frac{1}{\text{onset}_{i+1} - \text{onset}_i} \tag{2}$$

, where IBI_i represents the inter-beat-interval for the i-th beat. The unit of the tempo is beat-per-second.

Then, we translated the absolute value of the tempo into relative tempo ratio, such that

$$\text{tempo}_i' = \frac{\text{tempo}_i - \overline{\text{tempo}}}{\overline{\text{tempo}}} \tag{3}$$

For example, -0.2 means 20% slower than the average tempo, 0 means the same tempo as the average tempo, and

1 means doubling the average tempo. We limited this value to within -1 to 1.

During generation time, given the features of each beat, our system predicts the velocity and the relative tempo ratio of the beat. We then scaled the velocity to 1–127 and we calculated the onset time for each beat as

$$\text{onset}_i = \text{onset}_{i-1} + \frac{1}{\text{tempo}'_i \times \mathrm{T} + \mathrm{T}} \tag{4}$$

, where T is a constant that the user specifies the mean tempo to be. The unit of T is beat-per-second.

We encoded the generated performance in MIDI files capturing onset time, offset time, and velocity of each note. Due to the limitation of the dataset representation, the offset time of each note is unknown. Thus the prediction of the articulation (duration of the note in the performance over duration of the note in the score) is replaced with a fixed length. This does not affect our research about expressive timing since tempo is affected only by the onsets.

4.3 Training

We used 3-layer bi-directional LSTMs with 128 units to model beat-wise parameters. For velocity, the final layer is a sigmoid activation. For tempo prediction, the final layer is a tanh activation. The models were trained in a supervised fashion to minimize the mean-square-error loss. The sequence length was 64, the dropout rate was 0.5, and the learning rate was 0.001. We split the data into 80% training data and 20% validation data. The validation loss was 0.0412 for velocity and 0.0837 for tempo.

5. WHAT DOES THE NEURAL NETWORK LEARN?

5.1 Schemata

In section 2, we demonstrated that the characteristic signature rhythms are associated with the Mazurka Quality. To validate that the neural network can learn the Mazurka Quality on different signature rhythms, we encoded 16 measures of each signature rhythm as testing cases (the input scores) to feed into the beat-based neural network.

We then plotted the absolute tempo curve for each 16-bar signature rhythm input score. As in Figure 8, different input scores output different tempo curves. We see that in signature rhythms #2, #7, and #8, the output of the neural networks shows the "short-long" pattern, i.e., the tempo value of the first beat of each measure is higher than that of the second beat. Such pattern is especially strong on signature rhythm #8—there is on average a 26.6 beat-per-minute tempo difference between the first beat and the second beat. While in other rhythm as input, we see lengthening (slower tempo) of the first beat. This result aligns with Table 1, showing that the model learns about this general schema about Mazurkas.

5.2 Tempo Correlation

To evaluate the correlation of model-generated performance and human average performance, we calculated the Pearson's correlation coefficient (PCC) between computer-generated tempo and average human tempo for five pieces in the test set. From Table 2, we see that PCC-BH (PCC between the tempo generated by beat-based model and human average tempo) is higher than PCC-VH (PCC between the tempo generated by Virtuosonet [8] and human average tempo), indicating that the beat-based model learned generalized schematic performance practices more successfully than Virtuosonet in tempo estimation. As a reference, PCC between one human performance and other human performances is between 0.29 and 0.97.

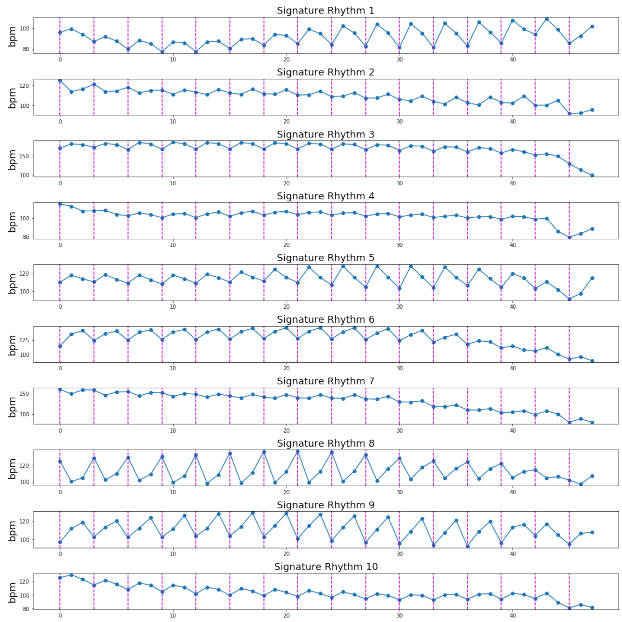


Figure 8: Tempo curves of different signature rhythm inputs. The "short-long" pattern is evident rhythms #2, #7, and #8.

Mazurka Op. #	PCC-VH	PCC-BH	PCC-HH
Op.17 No.4	0.065	0.151	0.794
Op.24 No.2	N/A	0.497	0.778
Op.30 No.2	0.048	0.44	0.786
Op.63 No.3	0.167	0.59	0.714
Op.68 No.3	0.489	0.59	0.889

Table 2: Pearson's correlation coefficient of performance tempo generated by Virtuosonet and human average tempo (PCC-VH), the beat-based model and human average tempo (PCC-BH), and a random human performance tempo and average human performance tempo (PCC-HH).

5.3 Veridical Events

When plotting the tempo curve of Op.63 No.3 (Figure 9), we see that the beat-based model (line 3) learns about the schema that when pattern #2 occurs, the downbeat gets shortened. This aligned mostly with human performances. For performances generated by Virtuosonet (line 4), since it is trained on 16 composers' pieces, it is understandable that it does not favor Mazurka's rhythmic tempo. Thus we see an almost opposite direction.

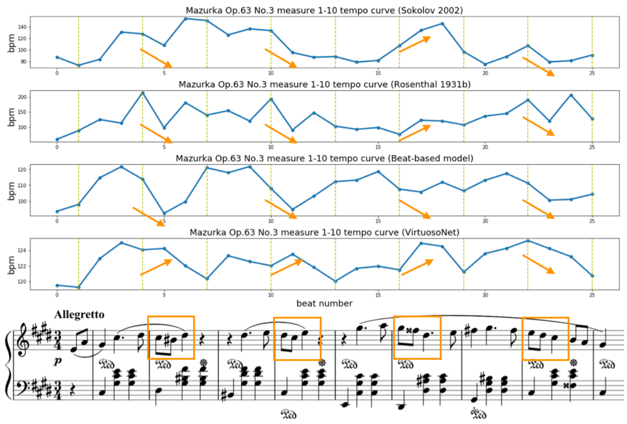


Figure 9: Tempo curves of the first 10 measures of Op.63 No.3 from two human performances (lines 1 & 2), a beat-based model (line 3) and Virtuosonet (line 4), with the score (line 5). Tempo trends of pattern #2 (boxed) are noted with arrows.

In an analysis of 95 distinct human performances of the same work, in beats 5 and 6 (the first rectangle in Figure 9), 89 performers shortened the downbeat, in beats 11 and 12 (the second rectangle), 82 performers shortened the downbeat, and in beats 23 and 24 (the fourth rectangle), 88 performers shortened the downbeat. However, in beats 17 and 18 (the third rectangle), there were more performers (53 out of 95) who lengthened the downbeat rather than shortening it. While this seems to be an “outlier”, we were interested to further investigate what’s happening musically on these two beats.

Harmonic analysis was performed on the first ten measures of the score as in Figure 10. Note that there is a secondary dominant in measure 7 (beats 17-19), whose veridical change of color and direction prompts performers to emphasize the moment. The beat-based network captured the schematic “short-long” accent across many Mazurka performances, however, performance of this salient veridical event was not compellingly captured.

Another example is the A phrase of Op.24 No.2 (as in Figure 11). This phrase consists of four sub-phrases. Each sub-phrase is two-bar long. When the machine played this passage, we saw a very consistent trend. As in Figure 12, for each sub-phrase of 2 bars (6 beats), the machine lengthened the second to last beat. In addition, for all four sub-phrases, the tempo curves were similar: during the first three beats the tempo surged, and for the next two beats the

Figure 10: Harmonic analysis of the first 10 measures of Op.63 No.3. The secondary dominant in measure 7 is a veridical characteristic not captured by the beat-based network.

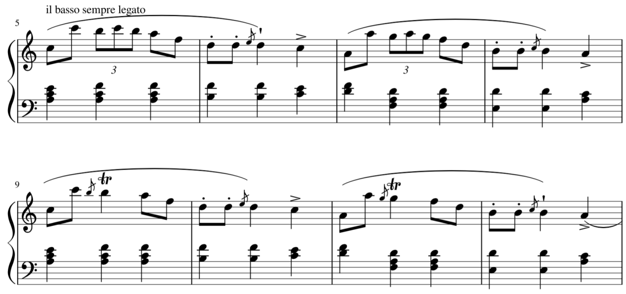


Figure 11: Phrase A of Op.24 No.2 . This phrase consists of four sub-phrases, each two-bar long.

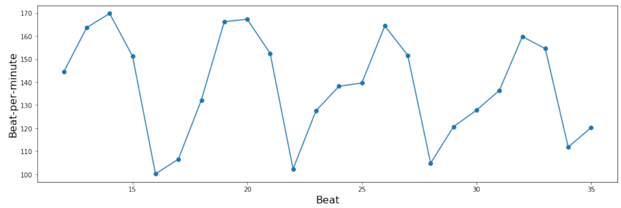


Figure 12: The tempo curve of machine playing the A Phrase of Op.24 No.2. For each sub-phrase, the machine emphasized (lengthened) the second to last note. This trend is consistent among all four sub-phrases.

tempo decreased, and finally the tempo slightly increased for beat 6. However, when we examined human performances, we found different results. As in Table 3, in contrast to the computer performance, human performers tend to vary the emphasized beat for each sub-phrase, whereas the computer performs the same one for each sub-phrase.

Sub-phrase	B1	B2	B3	B4	B5	B6
1	**24**	11	2	3	8	2
2	0	0	0	11	**27**	12
3	7	7	0	**26**	7	3
4	4	2	4	**25**	10	5

Table 3: Beats that human performers lengthened most for each sub-phrase in Op.24 No.2.

6. SUMMARY

In this paper we described our implementation of a beat-based model to learn expressive timing parameters in Chopin Mazurkas. Comparing human performances with performances generated by our model, we note that neural network succeeds at modeling schemas (such as distortion of the characteristic “short-long” Mazurka rhythm, and temporal augmentation at the approach to phrase-ends). However piece-specific veridical events (such as performed variations of repeated rhythmic units) are difficult to learn. One reason for this is that insufficient instances of examples of such veridical moments in the training set make it difficult for a deep learning-based system to acquire. Capturing the performance nuances of veridical events is a critical next step for the success of future computational models of expressive music performance.

7. REFERENCES

[1] B. Repp, "A microcosm of musical expression. i. quantitative analysis of pianists' timing in the initial measures of chopin's etude in e major," *The Journal of the Acoustical Society of America*, vol. 104, no. 2, pp. 1085–1100, 1998.

[2] C. Palmer, "Music performance," *Annual review of psychology*, vol. 48, no. 1, pp. 115–138, 1997.

[3] A. Friberg, R. Bresin, and J. Sundberg, "Overview of the kth rule system for musical performance," *Advances in Cognitive Psychology*, vol. 2, no. 2-3, pp. 145–161, 2006.

[4] G. Widmer, "Machine discoveries: A few simple, robust local expression principles," *Journal of New Music Research*, vol. 31, no. 1, pp. 37–50, 2002.

[5] M. Grachten and G. Widmer, "Linear basis models for prediction and analysis of musical expression," *Journal of New Music Research*, vol. 41, no. 4, pp. 311–322, 2012.

[6] C. E. Cancino-Chacón, T. Gadermaier, G. Widmer, and M. Grachten, "An evaluation of linear and non-linear models of expressive dynamics in classical piano and symphonic music," *Machine Learning*, vol. 106, no. 6, pp. 887–909, 2017.

[7] C. E. C. Chacón, "Computational modeling of expressive music performance with linear and non-linear basis function models," Ph.D. dissertation, Johannes Kepler University Linz, Austria, 2018.

[8] D. Jeong, T. Kwon, Y. Kim, K. Lee, and J. Nam, "Virtuosonet: A hierarchical rnn-based system for modeling expressive piano performance," in *Proceedings of the 20th International Society for Music Information Retrieval*, Delft, The Netherlands, 2019, pp. 908–915.

[9] A. Maezawa, K. Yamamoto, and T. Fujishima, "Rendering music performance with interpretation variations using conditional variational rnn," in *Proceedings of the 20th International Society for Music Information Retrieval*, Delft, The Netherlands, 2019, pp. 855–861.

[10] N. Todd, "A model of expressive timing in tonal music," *Music Perception: An Interdisciplinary Journal*, vol. 3, no. 1, pp. 33–57, 1985.

[11] E. Chew, "Playing with the edge: Tipping points and the role of tonality," *Music Perception: An Interdisciplinary Journal*, vol. 33, no. 3, pp. 344–366, 2016.

[12] J. Peperkamp, K. Hildebrandt, and C. C. S. Liem, "A formalization of relative local tempo variations in collections of performances," in *Proceedings of the 18th International Society for Music Information Retrieval Conference*, 2017, pp. 158–164.

[13] R. Gjerdingen, *Music in the Galant Style*. Oxford University Press, 2007.

[14] K. Kosta, O. Bandtlow, and E. Chew, "MazurkaBL: score-aligned loudness, beat, expressive markings data for 2000 chopin mazurka recordings," in *Proceedings of the fourth International Conference on Technologies for Music Notation and Representation (TENOR)(Montreal, QC)*, 2018, pp. 85–94.

[15] C. S. Sapp, "Comparative analysis of multiple musical performances." in *ISMIR*, 2007, pp. 497–500.

[16] A. Swartz, "The polish folk mazurka," *Studia musicologica Academiae Scientiarum Hungaricae*, vol. 17, no. Fasc. 1/4, pp. 249–255, 1975.

[17] J. J. Bharucha, "Tonality and expectation." *Musical perceptions*, pp. 213–239, 1994.

[18] M. Grachten, C. E. Chacón, and T. Gadermaier, "partitura: A python package for handling symbolic musical data," in *Late-Breaking Demo Session of the 20th International Society for Music Information Retrieval Conference. Delft, The Netherlands*, 2019.