# A demonstration of hierarchical structure usage in expressive timing analysis by model selection tests

Shengchen Li[1], Simon Dixon[2], Mark D. Plumbley[3]

1. Beijing University of Posts and Telecommunications, Beijing 100876, P. R. China
E-mail: shengchen.li@bupt.edu.cn

2. Queen Mary University of London, United Kingdom
E-mail: s.e.dixon@qmul.ac.uk

3. University of Surrey, United Kingdom
E-mail: m.plumbley@surrey.ac.uk

Abstract: Analysing expressive timing in performed music can help machine to perform various perceptual tasks such as identifying performers and understand music structures in classical music. A hierarchical structure is commonly used for expressive timing analysis. This paper provides a statistical demonstration to support the use of hierarchical structure in expressive timing analysis by presenting two groups of model selection tests. The first model selection test uses expressive timing to determine the location of music structure boundaries. The second model selection test is matching a piece of performance with the same performer playing another given piece. Comparing the results of model selection tests, the preferred hierarchical structures in these two model selection tests are not the same. While determining music structure boundaries demands a hierarchical structure with more levels in the expressive timing analysis, a hierarchical structure with less levels helps identifying the dedicated performer in most cases.

Key Words: Expressive timing, Hierarchical Analysis, Model Selection

## 1 Introduction

A certain number of music theories such as Generative Theory of Tonal Music (known as GTTM) [1] considers the music structure as a hierarchical structure. As music structure affects expressive timing, hierarchical structures are commonly used in expressive timing analysis. In this paper, expressive timing is analysed with different structures to provide a statistical evidence that hierarchical structures are helpful in the analysis of expressive timing.

In this paper, expressive timing is analysed with hierarchical models that assert probability of locating a music structure boundary on each beat of performed music. The resulting models are then used to detect music structure boundaries and identify the performances by the same performer of a given piece. To compare different structures used in expressive timing analysis, the proposed method is capable of different structures of analysis including both hierarchical and non-hierarchical structures with multiple levels.

There are several candidate structures of analysis are compared in this work. Besides the prime target comparing hierarchical and non-hierarchical structures, there are different number of levels in the candidate structures with the units in the basic level engaging to different number of beats. Understanding the preference of hierarchical structure for analysing the connections between expressive timing and different aspects of musicology help machines to understand music perceptually thus forms basis of various tasks such as performer identification and cover song identification.

To assert the probability of locating music structure boundaries, the hierarchical parabola model by Todd [2] is used inversely. By finding the local minima of expressive timing at different levels, the expressive timing information can be converted to a probabilistic model that each beat in a piece of expressive performance locates a boundary of music structure. The resulting models are then compared by Receiver Operation Characteristic (ROC) test. Considering that expressive timing is affected by music structure boundaries [2], the results of the model selection test reveals the preference of structures showing the commonalities in expressive performances.

As the way that a performer interprets music structure boundaries is possibly unchanged, the probability model asserting each beat in the performance locates a music structure boundary may be used to identify same-performer renderings. The term same-performer rendering means the performances of the same piece of music that is played by the same performer at different times. Assuming the unchanged mapping between expressive timing and the understanding of music structure, the resulting probability model locating music structure boundaries should be similar between same-performer renderings.

To show the proposed algorithm can be potentially applied to other classical music performances, the most commonly used expressive timing dataset in classical piano performances, the Mazurka dataset [1] is used. This database contains various performances of five pieces of Mazurkas (Op.17/4, Op.24/2, Op.30/2, Op.63/3 and Op.68/3) and is commonly used in various works [3–6]. For each piece of music, there are multiple perfor-

[1]http://www.mazurka.org.uk/info/excel/beat/

mances. Certain sets of same-performer renderings are also offered. Besides the information on expressive timing, the Mazurka database also provides the boundaries of the music structure in each Mazurka, which is used to evaluate the models resulting from the candidate structures.

The proposed method is not an empirical method. In other words, the proposed method does not require data from various performances but only needs expressive timing information from a single piece of performance. The principle of the proposed method is taken the expressive timing that is recorded as a vector of tempo variations as a signal. Using signal processing methods, the proposed method is capable for various structures of analysis.

As both music structure boundaries assertion and performer identification can be regarded as a binary classification problem, the performance of models resulting from different structures are evaluated by Receiver Operating Characteristic (ROC) [7]. The specific measure used to evaluate the model performance is Area Under Curve (AUC) [7], which represents how well the resulting model can perform compared with random selection results.

This paper is organised as follows: relevant research is reviewed first. The introduction of data used in this paper follows. Then the proposed method that converts expressive timing to the probability that each beat in a piece of performance locates a music structure boundary is explained in detail before the presentation of candidate structures for analysis. The final results are shown with a following discussion and conclusion.

## 2   Literature Review

In this paper, we build a model that converts expressive timing into the probability that every beat in a performance locates a music structure boundary. According to existing works [3, 8], the expressive timing and the phrasing of music are closely related; thus, an algorithm is proposed asserting the probability that each beat in a performance locates a music structure boundary according to the minima of expressive timing at different levels. The introduction of structure for analysis defines the local minima of expressive timing in different way. Thus the preferred hierarchical structure for analysing expressive timing can be selected by comparing different candidate structures for analysis.

There are two types of tasks for expressive timing analysis using a hierarchical structure. Finding the commonalities of expressive timing by different performers forms a basis of expressive timing generation. The classical parabola model [2] uses parabolic curves to regress expressive timing at different levels. With the same principle, earlier attempts by Desain and Honing [9] uses the hierarchical parabola model to synthesis a MIDI piece but have little success. Later, the introduction of machine learning techniques further improves the synthesis of expressive timing. The DISTALL system [10–12] considers the hierarchical relationship in tempo variations and use such relationship to synthesise expressive timing. As the expressive timing is affected by music structure and the music structure is hierarchical [1], the proposed experiment asserting music structure boundaries compares which structure is preferred for finding commonalities in expressive timing.

The hierarchical method is also used for characterising personal expressive style. Sapp [5] uses a multi-layer algorithm to identify performances from the same performer. With multiple lengths of windows, the personal expressive style is compared with the same piece of music performed. Moreover, Saunders et al. [13] and Molina-Solana et al. [14] have used expressive information to identify performers with reasonable results presented.

In model selection tests, the proposed algorithm of expressive timing analysis is performed with different structure of analysis adapted. To evaluate how well the candidate models perform, a ROC analysis [15] is performed to examine how well each candidate structure of analysis performs in the dedicated tasks. A ROC analysis balances the hit rate and the rate of false positive. Setting a threshold to judge the positives, it is likely to obtain two types of results: true positives and false positives. If we use the rates of true positives and false positives as coordinates, we can obtain a set of points on a plane. A curve connecting all points is known as the ROC curve. The area under the curve (known as AUC) indicates the efficiency of the tested model in the task. A larger AUC means better performance. We will compare a set of models resulting from a set of proposed structures.

## 3   Dataset

The term expressive timing is a perceptual concept that the beat of performed music varies to express different emotions. A straightforward way to digitise expressive timing is to use a vector of tempo variations.

Despite beat tracking algorithms reaches a reasonable level of accuracy by music information retrieve researchers, the accuracy of beat tracking for classical music is still far from usable due to the rapid changes of beat lengths. As a common practice, the raw data of expressive timing still demands human annotation works. Labelling the timing of each beat in expressive performance with a vector of beat timing $\mathbf{T} = (t_1, t_2, \ldots, t_n)$ where there are $n$ beats in the performance, the tempo of each beat can be calculated by the reciprocal of inter-beat interval, i.e. $\tau_i = \frac{1}{t_i - t_{i-1}}$. Connecting the values of tempo on each beat, the resulting curve is named as tempo curve that is a common way to visualise expressive timing.

Figure 1 compares the tempo curves throughout a piece of music for two different players performing the same piece of music. The vertical axis shows the values of tempo on each beat. The unit of beat is Beat Per Minute (BPM) where a larger number indicates a faster rate of beat. The horizontal axis is the index of beat.

In Figure 1, the tempo curves show a possible hierarchical structure: each tempo curve is formed by several arcs and the peak of each arc forms further arc shapes.
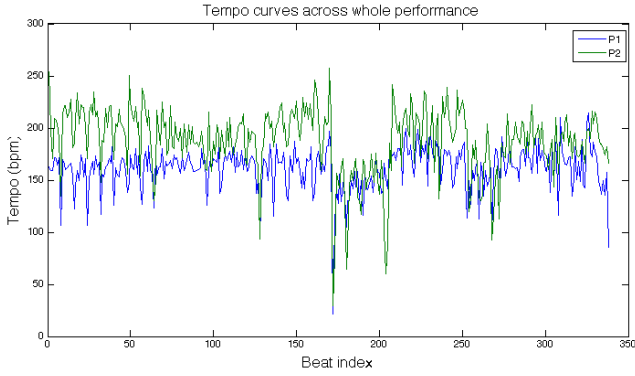
Fig. 1: Tempo variations of two different performers playing the same piece of music.



Fig. 2: Music structure analysis for phrasing of Chopin Mazurka Op.24 No.2.

Moreover, at certain positions, the performers agree on the change of tempo whereas for other positions the micro-structure of each arc differs from performer to performer. The agreed positions are usual the boundary of music structure.

The music structure analysis in this paper focuses on phrasing. Figure 2 shows an example of the result of music structure analysis. There are colour blocks cover the music sheet in Figure 2 labelling the phrases in a piece of music. The last beat of each phrase is defined as the music structure boundary. The piece shown in Figure 2 is Chopin Mazurka Op.24 No.2 that is included in the Mazurka dataset used in this paper.

Provided both the result of music structure analysis and expressive timing formation in terms of tempo variations, Mazurka dataset contains five pieces of Mazurkas played by certain numbers of players. Next we will introduce how the expressive timing is converted to a probability model that asserts each beat in a performance locates a music structure boundary.

## 4 Model establishment

The models established show the probability that each individual beat throughout a performance locates

a music structure boundary. The proposed experiment asserts whether a beat locates a music structure boundary by comparing the probability of a beat in the resulting model with a selected threshold. As each performance generates a model with a given structure, the similarity of resulting models can be measured and the most similar performances are likely to be performed by the same performer. As the primary task of this paper is to show different hierarchical structures for analysis are preferred by different tasks in different tasks of expressive timing analysis, a single model is used for both tasks for comparison purposes hence the methods proposed in this paper is far from the optimised models for the specific music information retrieval tasks.

According to Todd [2], there is a potential hierarchical relationship in the tempo variations in performed music. This fact suggests that the global minimum points on tempo variations have a higher possibility of locating a boundary of a music structure, whereas maxima of tempo variations are less likely to be located at a music structure boundary. In other words, performers usually slow down the tempo around the important boundaries of the music structure. We use this principle inversely; thus, a minima of expressive timing has a higher probability of locating a music structure boundary.

The proposed model aims to describe, if there is a music structure boundary, where the boundary is likely to be located. Suppose that here are $n$ beats in the performance ($B = b_1, b_2, \ldots, b_n$); the model ($p(B) = (p_1, p_2, \ldots, p_n)$, $\sum_{p=1}^{n} = 1$) indicates that the probability of a music structure boundary $\beta$ is located on beat $i$ is $p_i$. The structure for analysis indicates how many beats are considered in a window at multiple levels. The different size of windows is represented as $\mathbf{L} = \{l_0, l_1, l_2, \ldots, l_i, \ldots, l_L\}$, where $l_0 < l_1 < l_2 < \ldots < l_L$. Specifically, $l_0$ represents the level of the beat (i.e., $l_0 = 1$). We use rectangle windows in this algorithm, which define the window function ($W_{l_i}(\tau)$) of size $l_i$ as follows:

$$W_{l_i}(\tau) = \begin{cases} 1 & \tau \in [0, l_i - 1] \\ 0 & \text{otherwise.} \end{cases} \quad (1)$$

For simplicity, we call the analysis with windows of size $l_i$ the analysis of level $i$. Level 1 is defined as the lowest level, and level $L$ is defined as the top level (or the highest level).

The root of mean square (RMS) of tempo values in each window at different levels is then calculated. If the window size is $l_i$ and the performance has $n$ beats (from 1 to $n$), we should have $\frac{n}{l_i}$ windows. However, if $n$ is not divisible by $l_i$, the size of the last window at level $l$ will be shortened to fit the end of the performance. We use $\mathbf{e}_{l_i} = (e_{l_i}^1, e_{l_i}^2, \ldots, e_{l_i}^j, \ldots, e_{l_i}^k)$ (where $k = \lfloor \frac{n}{l_i} \rfloor$) to represent the RMS values within each window whose size is $l_i$ at level $i$. The element in $\mathbf{e}_{l_i}$ is defined as follows:

$$e_{l_i}^j = RMS(\vec{\tau} W_{l_i}(\tau - jl_i)) - std(\vec{\tau}^T W_{l_i}(\tau - jl_i)), \quad (2)$$

where $RMS$ represents the root of mean square and $std$

represents the standard deviation. We must point out that $\mathbf{e}_{l_i}$ has different lengths. The length of $\mathbf{e}_{l_i}$ is $\lfloor \frac{n}{l_i} \rfloor$.

Next, we introduce the method to convert a series of $\mathbf{e}$ at different levels into the probability of locating a music structure boundary for every beat in a performance. The target of the proposed model is to determine the probability of locating one music structure boundary $\beta$ throughout a performance. For simplicity, the probability of a music structure boundary $\beta$ being located on beat $i$ is represented as $p(i) \equiv p(\beta = b_i)$ and the probability of a music structure boundary $\beta$ being located in window $W_{l_i}^j$ is represented as $p(W_{l_i}^j) \equiv p(\beta \in W_{l_i}^j)$. In the proposed model, we assume that the probability of locating a music structure boundary on a beat ($b_i$) is the product of the probability of locating a music structure in the windows ($W_{l_i}^{j_i}$) at all levels that the beat ($b_i$) is in. Suppose that there are $L$ levels in the candidate hierarchical structure of analysis; we have the following:

$$p(b_i) = \prod_{i=0}^{L} p(W_{l_i}^{j_i}). \tag{3}$$

In the proposed algorithm, the probability of a window $W_{l_i}^j$ at level $i$ is proportional to the reciprocal of squared $\mathbf{e}_{l_i}$; i.e.,

$$p(W_{l_i}^j) = \frac{1/(\mathbf{e}_{l_i}^j)^2}{\sum_{x=1}^{\lfloor \frac{n}{l_i} \rfloor}(1/(\mathbf{e}_{l_i}^x)^2)}. \tag{4}$$

## 5   Candidate hierarchical structures

Several hierarchical structures ($\mathbf{L}_0$ to $\mathbf{L}_7$) are proposed for comparison in this paper. The symbol $\mathbf{L}$ represents a series of window sizes for analysis. To represent each beat in the performance, the lowest level ($l_0$) in $\mathbf{L}$ is always 1. As a result, the lowest adjustable level in the candidate hierarchical structures for analysis is $l_1$.

In Mazurkas, there are three beats in each bar. Thus in most cases (except only two cases for comparison purposes), the second level of $\mathbf{L}$ is set at 3 ($l_1 = 3$). We want to investigate the effects of the multiple levels used. We set higher levels in the candidate hierarchical structure at 2 to group pairs of units at the previous level. The largest window size we used in the candidate hierarchical structures is 48 beats, which exceeds the longest phrases in the Mazurka database. We test two special cases in the candidate hierarchical structures by setting different values at $l_1$ for comparison purposes. The first special hierarchical structure has $l_1 = 6$, which is two bars. We want to investigate the effects such that $l_1$ covers a longer section of a performance. The second special hierarchical structure engaged is $l_1 = 4$, which conflicts with the grouping implied by the time signature. We wish to investigate whether it is important to make $l_1$ identical to the bar length or at least consistent with it. The candidate hierarchical structures are listed in Table 1.

## 6   Model Evaluation

To test the effects of candidate hierarchical structures on the proposed algorithm, we design two tasks for the

| Candidate Hierarchical Structures |
|---|
| $L_0 = \{1\}$ |
| $L_1 = \{1, 3\}$ |
| $L_2 = \{1, 3, 6\}$ |
| $L_3 = \{1, 3, 6, 12\}$ |
| $L_4 = \{1, 3, 6, 12, 24\}$ |
| $L_5 = \{1, 3, 6, 12, 24, 48\}$ |
| $L_6 = \{1, 6, 12, 24\}$ |
| $L_7 = \{1, 4, 12, 24\}$ |

Table 1: The candidate hierarchical structures for analysis in this experiment.

models and perform a ROC analysis. The two tasks we used in this paper are identifying a boundary of a music structure and identifying the same-performer rendering. According to the proposed model ($p(B)$), we can assert the location of music structure boundaries according to the resulting models with a threshold. Then, we compare the detected music structure boundaries by the resulting models with the music structure boundaries provided by the Mazurka dataset. We then follow a typical procedure of ROC analysis and then compare the models resulting from different candidate structures of analysis.

Suppose that the music structure provided by the Mazurka dataset is represented by a binary vector $\mathbf{\Theta}$; the music structure detected by a threshold $\omega$ can be represented as follows:

$$\mathbf{\Theta}_\omega = (p(B) < \omega). \tag{5}$$

Suppose that there are $k$ thresholds used in ROC analysis ($\omega_1, \omega_2, \ldots, \omega_k$, where $\omega_1 < \omega_2 < \ldots < \omega_k$), thus, there will be $k$ points on the ROC curves. The coordinates of the points on ROC curve are as follows:

$$\begin{cases} x_j = \text{sum}(\mathbf{\Theta}_{\omega_j} \text{ AND } (\text{NOT } \mathbf{\Theta})) \\ y_j = \text{sum}(\mathbf{\Theta}_{\omega_j} \text{ AND } \mathbf{\Theta}). \end{cases} \tag{6}$$

The AUC can be calculated as follows:

$$\text{AUC} = \sum_{j=1}^{k-1}(x_{j+1} - x_j)\frac{y_{j+1} + y_j}{2} \tag{7}$$

where a larger AUC value indicates a better result.

In a previous work [5], the author uses a correlation-based similarity measurement to compare the performances from the same performer, or same-performer renderings. According to this work, same-performer renderings are more similar to each other than different-performer renderings. In the current paper, we investigate which structure of analysis helps show the similarity between same-performer renderings.

To show the similarity between same-performer renderings, we compare the model we proposed from all renderings, including the same-performer renderings and the different-performer renderings for the same Mazurka first. For a particular hierarchical structure, the similarity of resulting models is compared by KL divergence [16, p. 58].

| ROC/AUC | $L_0$ | $L_1$ | $L_2$ | $L_3$ | $L_4$ | $L_5$ | $L_6$ | $L_7$ |
|---------|-------|-------|-------|-------|-------|-------|-------|-------|
| Op.17/4 | 0.2504 | 0.2538 | 0.2470 | 0.2521 | 0.2638 | 0.2807 | 0.2525 | 0.2531 |
| Op.24/2 | 0.2990 | 0.3106 | 0.3324 | 0.3788 | 0.4024 | 0.4107 | 0.3748 | 0.3517 |
| Op.30/2 | 0.0977 | 0.1070 | 0.1232 | 0.1428 | 0.1560 | 0.1639 | 0.1272 | 0.1227 |
| Op.63/3 | 0.2566 | 0.2624 | 0.2878 | 0.3148 | 0.3178 | 0.3300 | 0.3038 | 0.2939 |
| Op.68/3 | 0.2507 | 0.2475 | 0.2440 | 0.2654 | 0.2970 | 0.3010 | 0.2891 | 0.2783 |
| Mean | 0.2309 | 0.2363 | 0.2469 | 0.2708 | 0.2874 | 0.2973 | 0.2695 | 0.2599 |

Table 2: The AUC value for identifying music structure boundaries. A larger number suggests that a better structure is used in the analysis. The bold number indicates the best structure of analysis for a Mazurka.

| ROC/AUC | $L_0$ | $L_1$ | $L_2$ | $L_3$ | $L_4$ | $L_5$ | $L_6$ | $L_7$ |
|---------|-------|-------|-------|-------|-------|-------|-------|-------|
| Op.17/4 | 0.9915 | 0.9936 | 0.9907 | 0.9697 | 0.9734 | 0.9657 | 0.9688 | 0.9937 |
| Op.24/2 | 0.9945 | 0.9953 | 0.9947 | 0.9926 | 0.9785 | 0.9680 | 0.9632 | 0.9760 |
| Op.30/2 | 0.9546 | 0.9604 | 0.9703 | 0.9710 | 0.9671 | 0.9230 | 0.9516 | 0.9720 |
| Op.63/3 | 0.9883 | 0.9925 | 0.9864 | 0.9844 | 0.9800 | 0.9708 | 0.9791 | 0.9812 |
| Op.68/3 | 0.9511 | 0.9552 | 0.9412 | 0.9200 | 0.9167 | 0.9102 | 0.9282 | 0.9267 |
| Mean | 0.9760 | 0.9794 | 0.9767 | 0.9675 | 0.9631 | 0.9475 | 0.9582 | 0.9699 |

Table 3: The AUC value for identifying same-performer renderings. A larger number suggests that a better structure is used in the analysis. The bold number indicates the best structure of analysis for a Mazurka.

Selecting a performance as target rendering, we use $D = (d_1, d_2, \ldots, d_m)$ to represent the KL divergence between the model derived by the target rendering and the models derived by other renderings. Again we choose $k$ thresholds $(\omega_1, \omega_2, \ldots, \omega_k$, where $\omega_1 < \omega_2 < \ldots < \omega_k)$. Thus the detected same-performer renderings can be represented by a binary vector $\Theta_{(\omega_j)}$, where

$$\mathbf{\Theta}_{\omega_j} = (D < \omega_j). \tag{8}$$

Given the ground truth ($\mathbf{\Theta}$) of the same-performer renderings as provided by the Mazurka database, equation (6) and (7) can be used again to calculate the AUC value.

## 7 Results and Discussion

In Table 2 and Table 3, the resulting AUC values in music structure boundary assertion and same-performer rendering identification are shown. In both tasks, we can see that, a hierarchical structure in the analysis is helpful for the analysis of expressive timing in general but the preferred structure of analysis differs in different tasks.

As shown in Table 2, a particular structure $L_5$ outperforms the other structures of analysis regardless of the piece of music used for analysis when the proposed method is used to identify music structure boundaries. Examining the structure of analysis $L_5$, the $L_5$ has the highest number of layers and that the top-level windows in $L_5$ cover the widest area. Comparing the AUC resulting from $L_1$ to $L_5$ more levels and wider coverage for the top-level window help the identification of music structure boundaries for the proposed methods. For structure $L_7$, which breaks the metric structure of Mazurkas, we find that the AUC results are worse than the structure that has the same number of layers ($L_3$) and the structure that covers the same length at the top level ($L_4$). This fact suggests that a structure breaking the metric structure of a piece of music may lead to worse performance of the analysis. Moreover, the hi-

erarchical structures ($L_1$ to $L_7$) can identify the music structure boundaries more precisely compared with the flat structure of analysis ($L_0$).

The results of identifying same-performer renderings are shown in Table 3, where the results in general are better than Molina-Solana et al. [14] and are the comparable with Saunders et al. [13] with reduced complexity. Among the results, two structures of analysis outperforms the other structure of analysis depending on different Mazurkas. For Op.17/4 and Op.30/2, the best model to identify same-performer renderings is $L_7$, which breaks the metric structure of Mazurka. In contrast, for all other Mazurkas, the less hierarchical structure $L_1$ helps the identification of same-performer renderings most. This fact suggests that the best hierarchical structure for identifying same-performer renderings differs across different Mazurkas. In general, unlike the case of identifying music structure boundaries, the structure of analysis for identifying same-performer renderings shows a preference for a flatter structure, such as $L_1$ and $L_2$. A structure of analysis with many layers, such as $L_4$, $L_5$, and $L_6$, do not help in identifying same-performer renderings; in contrast, these structures perform worse than the flat structure ($L_0$). This result suggests that only some hierarchical structures help in the identification of same-performer renderings. Moreover, as listed in Table 3, the value of AUC suggests that the proposed algorithm is fairly effective in identifying same-performer renderings in the Mazurka database.

In summary, the hierarchical structure in the analysis of expressive timing help to improve the performance of the proposed algorithm. However, for different tasks, the best structure for analysis is different. For analysing the mapping between music structure and expressive timing, a hierarchical structure with more levels are preferred. The hierarchical structure with fewer levels are preferred for identifying same-performer renderings. In some cases, the best structure for analysis depends on the specific pieces of music as well.

## 8 Conclusions

In this paper, a model selection test is performed to demonstrate the use of hierarchical in expressive timing analysis improves the performance of analysis. A method of converting expressive timing into the probability that each beat in the performance will locate a boundary of a music structure is proposed. This method is capable of adopting a hierarchical structure for analysis during the modelling process. There are a few different hierarchical structures used as inputs and then evaluated the resulting models by two different experiments.

To evaluate the model performances with different hierarchical structures, two model selection tests are performed. The first one used the resulting models to predict the boundaries of the music structure. The second test investigated how well the resulting models could find same-performer renderings. The way to evaluate the performance of the resulting models is ROC analysis.

Comparing the performance of several hierarchical structures in both tests, including a non-hierarchical structure and several hierarchical structures that keep the bars in the performances. Moreover, for comparison purposes, there is also a hierarchical structure that matches bar starts and ends in the performance but covers a larger area in the lower levels and a hierarchical structure that breaks bars in the performances.

Based on the results, a hierarchical structure in expressive timing analysis could be helpful. In detecting music structure boundaries, a more hierarchical structure with a wider coverage of top-level windows performs well. In identifying the same-performer renderings, we find that a structure with a certain level of hierarchy is helpful. Moreover, the most suitable structure of analysis for identifying same-performer renderings differs according to different pieces.

As a summary, the use of hierarchical structure in expressive timing analysis is demonstrated as helpful. With the tasks of identifying same-performer rendering and determining music structure boundaries, the hierarchical structure improves the model performance despite the different preference.

## References

[1] Fred Lerdahl and Ray Jackendoff. A Generative Theory of Tonal Music. MIT Press, 1983.

[2] Neil P. Mcangus Todd. The dynamics of dynamics: A model of musical expression. Journal of Acoustical Society of America, 91:3540–3550, 1992.

[3] Neta Spiro, Nicolas Gold, and John Rink. The form of performance: Analyzing pattern distribution in select recordings of Chopin's Mazurka op. 24 no. 2. Musicae Scientiae, 14(2):23–55, 2010.

[4] John Rink, Neta Spiro, and Nicolas Gold. Motive, gesture and the analysis of performance. New Perspectives on Music and Testure, pages 267 – 292, 2011.

[5] Craig Sapp. Hybrid numeric/rank similarity metrics for musical performance analysis. In Proceedings of the International Conference on Music Information Retrieval (ISMIR), pages 501–506, 2008.

[6] Shengchen Li, Simon Dixon, and Mark D. Plumbley. Clustering expressive timing with regressed polynomial coefficients demonstrated by a model selection test. In Proceedings of the 18th International Society of Music Information Retrieval Conference, Su Zhou, China, 2017.

[7] Christopher D. Brown and Herbert T. Davis. Receiver operating characteristics curves and related decision measures: A tutorial. Chemometrics and Intelligent Laboratory Systems, 80(1):24–38, jan 2006.

[8] Anders Friberg, Roberto Bresin, and Johan Sundberg. Overview of the KTH rule system for musical performence. Advances in Cognitive Psychology, 2:145–161, 2006.

[9] Peter Desain and Henkjan Honing. Tempo curves considered harmful. Contemporary Music Review, 7:123 – 138, 1993.

[10] Gerhard Widmer and Asmir Tobudic. Playing Mozart by analogy: Learning multi-level timing and dynamics strategies. Journnal of New Music Research, 32:259–268, 2003.

[11] Asmir Tobudic and Gerhard Widmer. Playing mozart phrase by phrase. In Proceedings of the 5th International Conference on Case-based Reasoning (ICCBR'03), pages 552–566. Springer, 2003.

[12] Asmir Tobudic and Gerhard Widmer. Relational ibl in music with a new structural similarity measure. In Proceedings of the 13th International Conference on Inductive Logic Programming (ILP'03), pages 365–382. Springer, 2003.

[13] Craig Saunders, David R. Hardoon, John Shawe-Taylor, and Gerhard Widmer. Using string kernels to identify famous performers from their playing style. intelligent data analysis, pages 384–395, 2004.

[14] Miguel Molinasolana, Josep Lluis Arcos, and Emilia Gomez. Using expressive trends for identifying violin performers. In Proceeding of International symposium on music information retrieval, pages 495–500, 2008.

[15] Tom Fawcett. An introduction to roc analysis. In Pattern Recognition Letters, volume 27, pages 861 – 874, 2005.

[16] Kevin P. Murphy. Machine Learning: A Probabilistic Perspective. The MIT Press, 2012.