CLUSTERING EXPRESSIVE TIMING WITH REGRESSED POLYNOMIAL COEFFICIENTS DEMONSTRATED BY A MODEL SELECTION TEST

Shengchen Li
Beijing University of
Posts and Telecommunications
shengchen.li@bupt.edu.cn

Simon Dixon
Queen Mary University of London
s.e.dixon@qmul.ac.uk

Mark D. Plumbley
University of Surrey
m.plumbley@surrey.ac.uk

ABSTRACT

Though many past works have tried to cluster expressive timing within a phrase, there have been few attempts to cluster features of expressive timing with constant dimensions regardless of phrase lengths. For example, used as a way to represent expressive timing, tempo curves can be regressed by a polynomial function such that the number of regressed polynomial coefficients remains constant with a given order regardless of phrase lengths. In this paper, clustering the regressed polynomial coefficients is proposed for expressive timing analysis. A model selection test is presented to compare Gaussian Mixture Models (GMMs) fitting regressed polynomial coefficients and fitting expressive timing directly. As there are no expected results of clustering expressive timing, the proposed method is demonstrated by how well the expressive timing are approximated by the centroids of GMMs. The results show that GMMs fitting the regressed polynomial coefficients outperform GMMs fitting expressive timing directly. This conclusion suggests that it is possible to use regressed polynomial coefficients to represent expressive timing within a phrase and cluster expressive timing within phrases of different lengths.

1. INTRODUCTION

In performed classical piano music, small variations of the beat length serving music expression is known as *expressive timing*. Expressive timing can be represented by *tempo curves* that connects the value of tempo on each beat to form a curve. A common method [3, 5, 8, 11] of analysing expressive timing within a phrase in performed classical piano music is to cluster expressive timing. One of the possible unit used for clustering expressive timing is *phrase* [5] that contains a certain beats forming a sensible music structure. The length of phrase, or *phrase length* (defined as the number of beats contained in a phrase), is expected to be identical throughout a piece of music by most algorithms such as Li et al. [5]. Such strong restrictions make the large-scale applications of existing algorithms almost impossible because the phrase lengths are not constant in most pieces. This paper proposes a way to cluster expressive timing regardless of phrase length.

In past research [14], polynomial functions, especially parabolic functions, are used to regress tempo curves. Regressing a tempo curve, the coefficients of the resulting polynomial function is called *regressed polynomial coefficients* for a tempo curve. Given an order of polynomial function to be regressed to, each tempo curve can be represented by a fixed number of regressed coefficients. In this paper, we propose to cluster regressed polynomial coefficients instead of clustering expressive timing directly in order to enable the clustering of expressive timing without a pre-defined unit possible. A model selection test is shown in this paper demonstrating the Gaussian Mixture Models (GMMs) fitting regressed polynomial coefficients outperform the GMMs fitting expressive timing directly.

For simplicity, the GMMs fitting the expressive timing are represented by GMM_o, whereas the GMMs fitting the regressed polynomial coefficients are represented as GMM_r. There are multiple ways to compare two GMMs. Because the two types of GMMs fitting two different sets of data in this paper, the traditional model selection criteria (such as Bayesian Information Criterion [1, Ch. 3] used by Li et al. [5]) based on model likelihood cannot be used. Although comparing the clustering results with a ground truth is a more general way to evaluate model performance, the clustering of expressive timing has no consensus or well-recognised "ground truth" by the musicologists. The performance of GMMs is evaluated by the approximation of each tempo curve by their corresponding centroids as this principle is a general evaluation for clustering algorithms.

To make the clustering of expressive timing and the regressed polynomial coefficients comparable, the pieces we selected in this paper still have constant phrase lengths. However, the expressive timing in various phrases can be regressed to the polynomial function of a single order regardless of phrase lengths. The three pieces of music are two pieces of Chopin's Mazurkas (Op. 24, No. 2 and Op. 30, No. 2) used in the previous works [10, 11] and Berekrev's *Islamey* dataset, which Li et al [5] used. Although the music analysed is classical music, the proposed algorithm for clustering expressive timing may be potentially used for other forms of music such as jazz music.

This paper is organised as follows: relevant literatures are reviewed first, then we describe how the standardised tempo curves and the regressed polynomial coefficients are clustered. Next, we will show how the performance of models are represented and the results are presented. A discussion comparing the differences of the GMMs precedes the conclusion of the paper.

2. BACKGROUNDS

Clustering is a widely used methodology for analysing expressive timing. As demonstrated by Li et al. [5], the standardised tempo curves within a phrase can be clustered. Repp [8] used Principle Component Analysis (PCA) to analyse expressive timing and found a certain number of common patterns. Spiro et al. [11] used self-organising maps to cluster expressive timing and expressive dynamics patterns within a bar and asserted that expressive timing and dynamics are affected by music structure. All these works requested a pre-selected unit of analysis with an identical length, such as bars, phrases or the entire piece of music. Such requirements, on the other hand, limit the usability of the methods of analysis because the choice of a unified unit for analysis is hard to find. Clustering regressed polynomial coefficients instead of expressive timing directly relaxes the restriction of a constant phrase length in the testing pieces; thus, more pieces of music can be analysed using different methodologies of clustering.

Using second-order polynomial function, or parabolic function, to regress expressive timing tempo curves representing expressive timing is a traditional way to model expressive timing [14]. This method was widely used in a range of past works [8, 9, 12, 15, 16]. Repp [8, 9] used PCA to analyse expressive dynamics and timing in certain numbers of performances of a Chopin's étude. Repp asserted that the parabolic curves are particularly good at modelling the expressive timing within a longer phrase unit [8] and that parabolic curves are useful for regressing the expressive dynamics [9]. Tobudic and Widmer [13] used multi-level parabolic curves to learn how a concert pianist varied both dynamics and tempo when playing several Mozart pieces. The learned methods were then used to automatically render performances of other pieces with success. Timmers [12] suggested that using parabolic curves to regress expressive parameters in performances is useful in vocal performances. Despite the wide usage of parabolic curves for modelling, it is rare to cluster expressive timing with the regressed parabolic coefficients or regressed polynomial coefficients. This paper intends to use a model selection test to demonstrate that regressed polynomial coefficients are a valid representation of expressive timing for clustering.

GMMs are used to fit the distribution of expressive timing within a phrase and regressed polynomial coefficients. The resulting GMM_o and GMM_r are compared in the proposed model selection test. A model selection test is a common method in machine learning research to test the fitness of data with a mathematical model [1, Ch. 1]. Li et al. [5] used this method to analyse expressive timing. Model selection tests were used to demonstrate expressive timing can be modelled by a clustered model [5] and to determine the factors that affect the selection of clusters of expressive timing [6]. Because GMM_o and GMM_r model two different datasets, the approximation of expressive timing by their corresponding centroids of GMM_o and GMM_o is used for evaluation in this paper.

We adapt the dataset used by Li et al. [5] in which each piece has a constant length phrase to make GMM_o and GMM_r comparable. The three testing pieces of music are Chopin's Mazurkas (Op. 24, No. 2 and Op. 30, No. 2) and *Islamey*, whose lengths of phrases are twelve beats, twenty-four beats and eight beats throughout the entire piece, respectively. For each testing piece, there are sixty-four, thirty-four and twenty-five performances.

In each performance, the timing of each beat is recorded as $\{t_1, t_2, \ldots\}$ and the tempo value on each beat τ_i can be calculated as the reciprocal of inter-beat interval, namely $\tau_i = \dfrac{1}{t_{i+1} - t_i}$. The tempo value is then smoothed by the method of moving window average with a window size of 3 (i.e. $\bar{\tau}_i = \dfrac{\tau_{i-1} + \tau_i + \tau_{i+1}}{3}$) to approximate human perception of tempo [2]. The expressive timing within a phrase can then be represented as a vector of tempi or tempo curve: $\mathcal{T} = \{\bar{\tau}_1, \bar{\tau}_2, \ldots, \bar{\tau}_n\}$, where n represents the total number of beats. The resulting standardised tempo curves $\mathbf{T}$ are obtained by setting the mean of each tempo curve to 1, e.g. $\mathbf{T} = \{\hat{\tau}_1, \hat{\tau}_2, \ldots, \hat{\tau}_n\}$, where $\hat{\tau}_i = \dfrac{n\bar{\tau}_i}{\sum_{j=1}^{n} \bar{\tau}_j}$. After the standardisation process, in each data set there are m samples of n-dimensional data to be clustered, where m represents the number of phrases in the testing piece of music. These samples are the raw data for clustering and regression.

3. MODEL EVALUATION

In this paper, a method to cluster expressive timing regardless of the length of phrase is proposed. As there are no musicological ground truth available, the candidate models are evaluated by a traditional way to assess unsupervised machine learning algorithms: how well the original data can be approximated by the centroids of clusters.

Before discussing how GMM_o and GMM_r are compared in details, we will firstly brief how the GMMs are trained to fit data with n dimensions. A traditional way to train a GMM distribution is to use the Expectation Maximisation (EM) algorithm [7, Ch. 11] which attempts to raise the model likelihood of the training data by adjusting the parameters in GMM. Particularly in this paper, all the GMMs are trained for ten times with random initialisation and the best GMM is selected as the resulting GMM.

Next we will show how GMM_o and GMM_r are compared. As these two types of GMMs fit different data, the traditional measurements based on model likelihood such as BIC (Bayesian Information Criterion) are not valid. As a result, we evaluate how well the expressive timing within a phrase is approximated by the centroids of the resulting GMM_o and GMM_r. We will discuss how the approximation is measured in this section.

3.1 Evaluation of the Clustered Standardised Tempo Curves

Suppose the expressive timing in the ith phrase can be represented as $\mathbf{T_i} = \{\hat{\tau}_1, \hat{\tau}_2, \ldots, \hat{\tau}_n\}$. The distribution of $\mathbf{T_i}$ can be fitted to an A-component GMM (GMM_o) as [5]

$$p(\mathbf{T_i}) = \sum_{k=1}^{A} \pi_k \mathcal{N}(\mathbf{T_i}|\vec{\mu}_k, \mathbf{\Sigma}_k). \quad (1)$$

The centroids of the resulting GMM_o can be represented as $\vec{\mu}_1, \vec{\mu}_2, \ldots, \vec{\mu}_A$. If $\vec{\mu_i} = (\mu_1, \mu_2, \ldots, \mu_n)$ is used to represent the centroids of the cluster that tempo curve of the ith phrase ($\mathbf{T_i} = \{\bar{\tau}_1, \bar{\tau}_2, \ldots, \bar{\tau}_n\}$) belongs to (where $\vec{\mu_i} \in \{\vec{\mu}_1, \vec{\mu}_2, \ldots, \vec{\mu}_A\}$), the correlation coefficient (ρ) and Euclidean distance ($\mathcal{D}$) between the corresponding centroids μ_i and the expressive timing within a phrase $\mathbf{T_i}$ are given by

$$\rho(\mathbf{T_i}, \vec{\mu_i}) = \frac{\sum_{j=1}^{n}(\bar{\tau}_j - \overline{\mathbf{T_i}})(\mu_j - \overline{\vec{\mu_i}})}{\sqrt{\sum_{j=1}^{n}(\bar{\tau}_j - \overline{\mathbf{T_i}})^2}\sqrt{\sum_{j=1}^{n}(\mu_j - \overline{\vec{\mu_i}})^2}} \quad (2)$$

$$\mathcal{D}(\mathbf{T_i}, \vec{\mu_i}) = \sqrt{\sum_{j=1}^{n}(\bar{\tau}_j - \mu_j)^2} \quad (3)$$

where $\overline{\mathbf{T_i}} = \frac{1}{n}\sum_{k=1}^{n}\hat{\tau}_k$ and $\overline{\vec{\mu_i}} = \frac{1}{n}\sum_{k=1}^{n}\mu_k$.

3.2 Evaluation of the Regressed Polynomial Coefficients

With the least square algorithm, the standardised tempo curves can be regressed to a oth order polynomial function $f^o(x) = \sum_{i=0}^{o} b_i x^i$. Thus the standardised tempo curve representing expressive timing in phrase i ($\mathbf{T_i} = \{\bar{\tau}_1, \bar{\tau}_2, \ldots, \bar{\tau}_n\}$) can be represented by a vector of regressed polynomial coefficients $\vec{\mathbf{B}}_i = (b_0, b_1, b_2, \ldots, b_o)$. A GMM ($\text{GMM}_g$) fitting the oth order of polynomial coefficients is represented by GMM_g^o. For clarity, GMM_g represents the GMMs fitting the regressed parabolic coefficient of any orders.

To prevent overfitting (e.g. the function used for regression is too complex to generalise the distribution of data), the order of polynomial function o should be smaller than the length of phrase n ($o < n$). The GMM_g^o can be trained to fit the distribution of $\vec{\mathbf{B}}_i$ such that

$$p(\vec{\mathbf{B}}_i) = \sum_{k=1}^{A} \pi_k \mathcal{N}(\vec{\mathbf{B}}_i|\vec{m}_k, \mathbf{\Sigma}_k). \quad (4)$$

If the expressive timing of phrase i ($\mathbf{T_i}$) belongs to cluster k whose centroid can be represented as $\vec{m}_k = (b_{m0}, b_{m1}, \ldots, b_{mo})$, the regressed polynomial curve of the expressive timing within phrase i ($\mathbf{T_i}$) can be represented as $f^o(x|\vec{m}_i) = \sum_{j=0}^{o} b_{mj}x^j = (x_1, x_2, \ldots, x_n)$. Thus the correlation coefficient (ρ) and Euclidean distance ($\mathcal{D}$) between the regressed polynomial curve $f^o(x|\vec{m}_i)$ and the expressive timing $\mathbf{T_i}$ are given by

$$\rho(\mathbf{T_i}, f^o(x|\vec{m}_i)) = \frac{\sum_{j=1}^{n}(\bar{\tau}_j - \overline{\mathbf{T_i}})(f_j - \overline{f^o(x|\vec{m}_i)})}{\sqrt{\sum_{j=1}^{n}(\bar{\tau}_j - \overline{\mathbf{T_i}})^2}\sqrt{\sum_{j=1}^{n}(f_j - \overline{f^o(x|\vec{m}_i)})^2}} \quad (5)$$

$$\mathcal{D}(\mathbf{T_i}, f^o(x|\vec{m}_i)) = \sqrt{\sum_{j=1}^{n}(\bar{\tau}_j - x_j)^2} \quad (6)$$

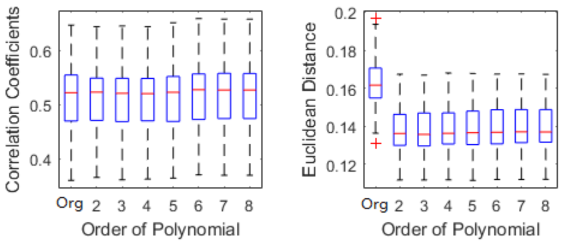
where $\overline{\mathbf{T_i}} = \frac{1}{n}\sum_{k=1}^{n}\hat{\tau}_k$ and $\overline{f^o(x|\vec{m}_i)} = \frac{1}{n}\sum_{k=1}^{n}x_k$.

4. RESULTS

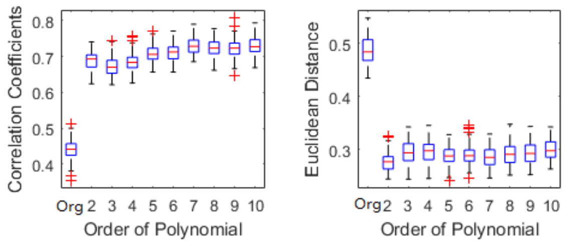
In this section, we will compare how the centroids of GMM_o and GMM_g approximate expressive timing within a phrase by showing the correlation coefficients and Euclidean distance discussed in Section 3. To train a GMM with the dataset selected, an important parameter should be decided: the intended number of clusters. Following the detailed discussion by Li et al. [5], we train GMMs with two Gaussian components for *Islamey*, eight Gaussian components for Chopin Mazurka Op.24/2 and four Gaussian components for Chopin Mazurka Op.30/2. Moreover, the order of polynomial function for regression is chosen between the second order and the tenth order for both Chopin's Mazurkas, whereas for *Islamey* whose phrase length is 8 beats the chosen order of polynomial function is between second order and the eighth order to prevent overfitting.

Compared with the complexity of the proposed GMM_o and GMM_g, the data we have is fairly limited. To prevent overfitting, cross validation is used in this experiment. Rather than using the entire dataset to train the GMM_o and GMM_g, only four-fifths of the performances form a training dataset, and the remaining performances form a testing dataset. Specifically for the candidate pieces, there are 5, 13, 7 performances used for testing and the numbers of performances for training are 20, 51, 27 for the candidate pieces *Islamey*, Chopin's Mazurka Op.24/2 and Chopin's Mazurka Op.30/2 respectively. To even out the possible bias caused by the randomness of the formation of the testing and training sets, cross validation tests are repeated for 100 times with the performances in the testing and training sets randomly selected. The performance of candidate models are evaluated by the average performance in the 100 cross validation tests.

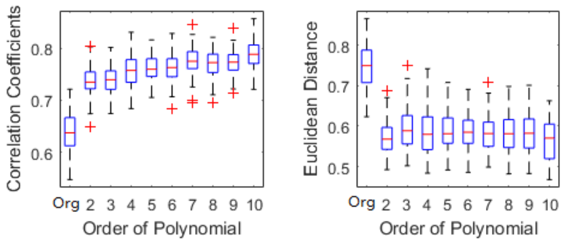
With the EM algorithm, a GMM_o and a GMM_g are trained with each training dataset engaged. The resulting GMM_o and GMM_g are used to cluster the testing dataset. The centroids of the resulting clusters are used to calculate $\rho(\mathbf{T_i}, \vec{\mu_i})$, $\mathcal{D}(\mathbf{T_i}, \vec{\mu_i})$, $\rho(\mathbf{T_i}, f^o(x|\vec{m}_i))$ and $\mathcal{D}(\mathbf{T_i}, f^o(x|\vec{m}_i))$ according to equations (2), (3), (5) and (6) where $\mathbf{T_i}$ is in the testing dataset. To remove the possible bias caused by the randomness of performance selection, the experiment is repeated 100 times. The resulting $\rho(\mathbf{T_i}, \vec{\mu_i})$, $\rho(\mathbf{T_i}, f^o(x|\vec{m}_i))$, $\mathcal{D}(\mathbf{T_i}, \vec{\mu_i})$ and $\mathcal{D}(\mathbf{T_i}, f^o(x|\vec{m}_i))$ are compared pairwisely.



(a) *Islamey*



(b) Mazurka Op.24/2



(c) Mazurka Op.30/2

Figure 1. Box plots of the resulting correlation coefficients and Euclidean distance between the standardised tempo curves and their corresponding cluster centroids. The box shows 25th and 75th percentiles. The line in the box shows the mean. Outliers are shown by a '+' sign. A higher correlation coefficients and a smaller Euclidean distance indicates a better approximation of expressive timing by corresponding centroids. The label 'Org' represents the results of clustering expressive timing directly.

In Figure 1, box plots of the resulting $\rho(\mathbf{T_i}, \vec{\mu}_i)$, $\mathcal{D}(\mathbf{T_i}, \vec{\mu}_i)$, $\rho(\mathbf{T_i}, f^o(x|\vec{m}_i))$ and $\mathcal{D}(\mathbf{T_i}, f^o(x|\vec{m}_i))$ in the 100 cross-validation tests for each testing piece are shown. In the diagram, the label 'Org' represents the performance of the centroids in GMM$_o$ (namely $\rho(\mathbf{T_i}, \vec{\mu}_i)$ and $\mathcal{D}(\mathbf{T_i}, \vec{\mu}_i)$,). The numbered labels represent the value of o in $\rho(\mathbf{T_i}, f^o(x|\vec{m}_i))$ and $\mathcal{D}(\mathbf{T_i}, f^o(x|\vec{m}_i))$. In each boxing plot, the box indicates the 25th and 75th percentiles and the line in the box shows the mean. The '+' signs show the outliers. A higher correlation coefficient and a smaller Euclidean distance means better approximation of the expressive timing by the corresponding centroids in GMM$_o$ and GMM$_g$.

In the resulting diagram, GMM$_g^o$ outperforms GMM$_g$ regardless of the value of o. As seen in Figure 1(b) and Figure 1(c), GMM$_g^o$ outperforms GMM$_r$ according to both the correlation coefficients and Euclidean distance. In Figure 1(a), although the correlation coefficients does not show that GMM$_g^o$ is better than GMM$_r$, the Euclidean distance shows that GMM$_g^o$ outperforms GMM$_r$. This result confirms that using polynomial functions to regress expressive timing within a phrase can help to improve the performance of clustering expressive timing.

Next, we discuss about which value of o makes the best performed GMM$_g^o$. With an one-way ANOVA test [7, Ch.8], we find that a higher value of o does not always introduce a better performance of GMM$_g^o$. To show the significance of the difference between the means and correlation coefficients of GMM$_g^o$ and GMM$_r$, we perform a Tukey's Honest Significant Difference (HSD) test.

With a preference of a simpler model, the results of Tukey's HSD show the following facts. If two GMM$_g^o$ have different values of o but no differences of performance, the GMM$_g^o$ with lower o values will be preferred due to lower complexity of GMM$_g^o$. For *Islamey*, the performance of GMM$_g^2$ to GMM$_g^8$ does not make significant differences thus GMM$_g^2$ is preferred. As a result, the second order of polynomial function is the most suitable method regressing expressive timing in *Islamey*. For Chopin Mazurka Op.24/2, GMM$_g^7$ to GMM$_g^{10}$ make no significant differences according to correlation coefficients whereas according to Euclidean distance, GMM$_g^{10}$ is worse than GMM$_g^7$ to GMM$_g^9$. So in general GMM$_g^7$ is the best model amongst the candidate models and the seventh order of polynomial function is the best function to regress expressive timing within a phrase for Mazurka Op.24/2. For Chopin Mazurka Op.30/2, the best performed models are GMM$_g^7$ to GMM$_g^{10}$ according to Euclidean distance whereas GMM$_g^{10}$ is marginally better than other models according to correlation coefficients. As a result, amongst the candidate models, the tenth order of polynomial function is the best model to regress the expressive timing within a phrase.

Considering the fact the phrase length of *Islamey*, Mazurka Op.24/2 and Mazurka Op.30/2 are 8 beats, 12 beats and 24 beats respectively and the most suitable polynomial function to regress expressive timing within a phrase is the second, the seventh and the tenth order, there may be a potential relationship between the most suitable order of polynomial function for regression and the phrase length. Demonstrating this hypothesis is beyond the scope of this paper but is possibly a future work.

5. DISCUSSION

5.1 Centroid Pairing

From the results of the model selection test, GMM$_r$ outperforms GMM$_o$. However, the resulting GMM$_r$ may not be necessary to make musical sense. As GMM$_o$ makes musical sense [4], the centroids of GMM$_r$ and GMM$_o$ are compared. If the regressed polynomial curves recovered from GMM$_r$ are correlated with the centroids of GMM$_o$, the GMM$_r$ will also make musical sense.

Recall that in Section 3, $\vec{\mu}_i$ represented the centroids of the GMMs for expressive timing within a phrase and $\vec{m}_j$ represented the centroids of the GMMs for regressed polynomial coefficients that can be recovered as a polynomial curve $f^o(x|\vec{m}_j)$. The similarity between centroids can be defined by the correlation coefficients (ρ) between $\vec{\mu}_i$ and $f^o(x|\vec{m}_j)$, namely

$$\rho(\vec{\mu}_i, f^o(x|\vec{m}_j)) = \frac{\sum_{k=1}^{n}(\mu_k - \overline{\vec{\mu}_i})(x_k - \overline{f^o(x|\vec{m}_j)})}{\sqrt{\sum_{k=1}^{n}(\mu_k - \overline{\vec{\mu}_i})^2}\sqrt{\sum_{k=1}^{n}(x_k - \overline{f^o(x|\vec{m}_j)})^2}} \quad (7)$$

where $\overline{\vec{\mu}_i} = \frac{1}{n}\sum_{k=1}^{n}\mu_k$ and $\overline{f^o(x|\vec{m}_i)} = \frac{1}{n}\sum_{k=1}^{n}x_k$.

Suppose there are A Gaussian components in the GMM_r and GMM_o. The centroids in GMM_r and GMM_o can be paired according to Algorithm 1. In Table 1, the results of pairing the centroids of the GMMs for the *o*th order polynomial coefficients are shown. In Figure 2, we demonstrate how the centroids of GMM_o compared with the regressed polynomial curves with the centroids of GMM_r^2 and GMM_r^7 in Chopin's Mazurka Op.24 No.2. From the results we can see that the regressed polynomial curves recovered from the centroids of GMM_r^2 are highly correlated with GMM_o while the regressed polynomial curves recovered from the centroids of GMM_r^7 are even more similar to GMM_o due to the higher model complexity. Thus the results demonstrate that the GMMs for the regressed polynomial coefficients are musically valid.

Algorithm 1 Pair centroids

Require: $f^o(x|\vec{m})_i$,$i \in [1, A]$
Require: $\vec{\mu}_j$,$i \in [1, A]$
  $\mathbf{C}(i, j) = \rho(\vec{\mu}_i, f^o(x|\vec{m}_j))$
  while $\max(\mathbf{C}) \geq -1$ **do**
    (r,c)=arg max ($\mathbf{C}_{rc}$)
    pairs:=pairs ∪ {r,c}
    Associate $\mathbf{loc}_r$ with $\mathbf{loc}_c$
    $\mathbf{C}_{ri}$ = -2
    $\mathbf{C}_{jc}$ = -2
  end while

	Islamey	Op.24/2	Op.30/2
GMM_r^2	0.99	0.83	0.75
GMM_r^3	0.99	0.82	0.74
GMM_r^4	0.99	0.84	0.72
GMM_r^5	0.99	0.89	0.80
GMM_r^6	1.00	0.90	0.78
GMM_r^7	1.00	0.90	0.82
GMM_r^8	1.00	0.90	0.76
GMM_r^9	N/A	0.90	0.79
GMM_r^{10}	N/A	0.90	0.86

Table 1. The correlation coefficients between the polynomial curves recovered from the centroids of GMM_r^o and the centroids of GMM_o.

5.2 GMM_r with more clusters

With the results presented, we can conclude that with the same Gaussian components in the model, GMM_g outperforms GMM_o when the intended number of Gaussian components is decided by the GMM_o provided by Li et al. [5]. In this section, we observe whether GMM_r, which has more Gaussian components, has a better performance. As an example, we compare the performance of GMM_r^2, measured by correlation coefficients with multiple Gaussian components. In Table 2, we show how well the regressed parabolic curves approximate the centroids of GMM_g^2 by showing $\rho(\mathbf{T_i}, f^o(x|\vec{m}_i))$ calculated by equation (5).

Clusters	*Islamey*	Op.24/2	Op.30/2
2	*0.5315*	0.5872	0.7339
4	0.5411	0.6181	*0.7399*
8	0.5718	*0.6877*	0.7442
16	0.6261	0.6930	0.7635
32	0.6722	0.7165	**0.7677**
64	0.6857	**0.7324**	0.7585
128	**0.6978**	0.7298	0.7413
256	0.6940	0.7282	N/A
512	0.6467	0.7109	N/A

Table 2. The average value of $\rho(\mathbf{T_i}, f^o(x|\vec{m}_i))$ resulting from GMM_g^2 with different numbers of Gaussian components (labelled as clusters in the table). A larger number means a better approximation and a better performance (bold). The number of clusters we set in the previous experiments are in italics. The training set of Mazurka Op.30/2 has less than 256 samples because it is impossible to set 256 and 512 clusters in the experiments.

From the table, we can see that the numbers of Gaussian components we engaged in the experiments in section 4 for GMM_g^2 do not have the best performance. Thus the GMM_g^2 with more Gaussian components can improve the model performance further.

6. CONCLUSIONS

In this paper, we demonstrate whether regressing standardised tempo curves within a phrase by a polynomial function is a valid method to analyse expressive timing by comparing Gaussian Mixture Models (GMMs) fitting expressive timing (GMM_o) and fitting regressed polynomial coefficients (GMM_g). As the candidate models fit different sets of data and there are no musicological ground truth for the clustering of expressive timing, the approximation of expressive timing by the centroids of GMM_o and GMM_r is used to evaluate model performance.

Measured by correlation coefficients and Euclidean distance, the experiment shows that GMM_g outperforms GMM_r when the same numbers of Gaussian components are engaged. With more Gaussian components engaged, GMM_r performs even better. The distribution of regressed polynomial coefficients has a lower degree of freedom compared with the tempo curves representing expressive timing within a phrase hence the regression of expressive timing with polynomial function reduces data dimension. The results demonstrate that regressing expressive timing with polynomial functions may help the clustering process.

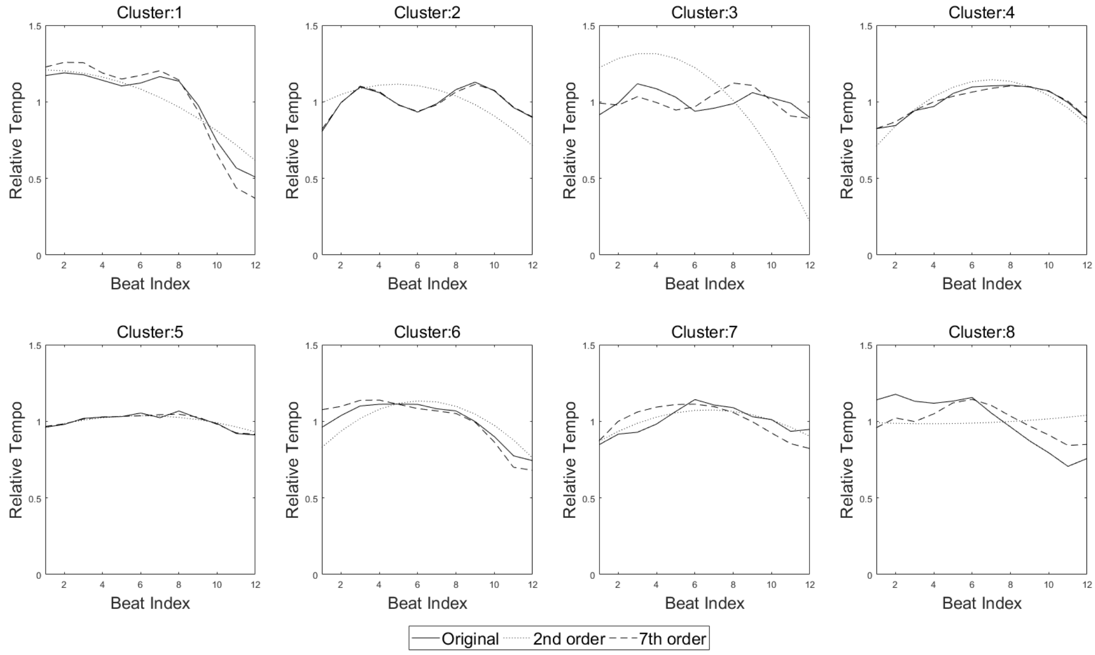


Figure 2. The centroids of GMM_o compared with the regressed polynomial curves with the centroids of GMM_r^2 and GMM_r^7 in Chopin's Mazurka Op.24 No.2.

When comparing the regressed polynomial curves recovered from the centroids of GMM_g with the centroids of GMM_o, the two sets of centroids are highly correlated with each other, which demonstrates that the centroids of GMM_g make similar musical sense with GMM_o. As a result, the polynomial functions can be used to help cluster expressive timing, which makes clustering expressive timing across phrases with various lengths possible.

7. ACKNOWLEDGEMENT

As a study based on my PhD research, we would like to thank Dr. Dawn A. A. Black and Prof. Elaine Chew for initial discussions. Chinese Scholarship Council partially founded the initial stage of the research.

8. REFERENCES

[1] Kenneth P. Burnham and David R. Anderson. *Model Selection and Multimodel Inference — A Practical Information-Theoretic Approach*. Springer, 2nd edition, 2002.

[2] Emilios Cambouropoulos, Simon Dixon, Werner Goebl, and Gerhard Widmer. Human preferences for tempo smoothness. In *Proceedings of the VII International Symposium on Systematic and Comparative Musicology and III International Conference on Cognitive Musicology*, pages 18–26, 2001.

[3] Maarten Grachten, Werner Goebl, Sebastian Flossmann, and Gerhard Widmer. Phase-plane representation and visualization of gestural structure in expressive timing. *Journal of New Music Research*, 38(2):183–195, 2009.

[4] Shengchen Li, Dawn A. A. Black, Elaine Chew, and Mark D. Plumbley. Evidence that phrase-level tempo variation may be represented using a limited dictionary. In *Proceedings of International Conference on Music Perception and Cognition (ICMPC'14)*, 2014.

[5] Shengchen Li, Dawn A. A. Black, and Mark D. Plumbley. The clustering of expressive timing within a phrase in classical piano performances by Gaussian mixture models. *Lecture Notes in Computer Science, Post Conference Proceeding of International Symposium on Computer Music Multidisciplinary Research*, 9617:322–345, 2016.

[6] Shengchen Li, Simon Dixon, Dawn A. A. Black, and Mark D. Plumbley. A model selection test for factors affecting the choice of expressive timing clusters for a phrase. In *Proceedings of 13th Sound and Music Computing Conference*, 2016.

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

[8] Bruno H. Repp. A microcosm of musical expression. I. Quantitave analysis of pianists' timing in the initial measures of Chopin's Etude in E major. *The Journal of the Acoustical Society of America*, 104:1085–1100, 1998.

[9] Bruno H. Repp. A microcosm of musical expression: II. Quantitative analysis of pianists' dynamics in the initial measures of Chopin's Etude in E major. *The Journal of the Acoustical Society of America*, 105:1972–1988, 1999.

[10] 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.

[11] 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.

[12] Renee Timmers. Vocal expression in recorded performances of Schubert songs. *Musicae Scientiae*, 11(2):237–268, 2007.

[13] 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.

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

[15] Gerhard Widmer. Discovering simple rules in complex data: A meta-learning algorithm and some surprising musical discoveries. *Artificial Intelligence*, 146:129–148, 2003.

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