EXPLAINING MUSICAL EXPRESSION AS A MIXTURE OF BASIS FUNCTIONS

Maarten Grachten
Department of Computational Perception
Johannes Kepler University, Linz, Austria
http://www.cp.jku.at/people/grachten

Gerhard Widmer
Department of Computational Perception
Johannes Kepler University, Linz, Austria
http://www.cp.jku.at/people/widmer

Austrian Research Institute for
Artificial Intelligence, Vienna, Austria

ABSTRACT

The quest for understanding how pianists interpret notated music to turn it into a lively musical experience, has led to numerous models of musical expression. One of the major dimensions of musical expression is loudness. Several models exist that explain loudness variations over the course of a performance, in terms of for example phrase structure, or musical accent. Often however, especially in piano music from the romantic period, performance directives are written explicitly in the score to guide performers. It is to be expected that such directives can explain a large part of the loudness variations. In this paper, we present a method to model the influence of notated loudness directives on loudness in piano performances, based on least squares fitting of a set of basis functions. We demonstrate that the linear basis model approach is general enough to allow for incorporating arbitrary musical features. In particular, we show that by including notated pitch in addition to loudness directives, the model also accounts for loudness effects in relation to voice-leading.

1. INTRODUCTION AND RELATED WORK

When a musician performs a piece of notated music, the performed music typically shows large variations in tempo, loudness, articulation, and, depending on the nature of the instrument, other dimensions such as timbre and note attack. It is generally acknowledged that one of the primary goals of such variations is to convey an expressive interpretation of the music to the listener. This interpretation may contain emotional elements (e.g. to play a piece 'solemnly'), and also elements that convey musical structure (e.g. to highlight a particular melodic voice, or to mark a phrase boundary) [1, 2].

These insights, which have grown over decades of music performance research, have led to numerous models of musical expression. The aim of these models is to explain the variations of loudness and tempo as a function of the structural interpretation of the music. For example, Todd [3] proposes a model of loudness that is a function of the phrase structure of the piece. Another example is Parncutt's model of musical accent [4].

Our current approach is limited to expressive dynamics. For this reason we will not discuss models of expressive timing here. More specifically, we will focus on the piano music of Chopin. This music is exemplary of classical music from the romantic period, which mainly evolved in Europe during the 19th century. Although this focus is admittedly very specific, it is often used to study expressive music performance (as in the seminal works of Repp [5]), since the music from the romantic period is characterized by dramatic fluctuations of tempo and dynamics.

Common dynamics annotations include *forte* (*f*), indicating a loud passage, *piano* (*p*) indicating a soft passage, *crescendo/decrescendo* indicating a gradual increase (resp. decrease) in loudness, respectively. Other, less well-known markings prescribe a dynamic evolution in the form of a metaphor, such as *calando* ("growing silent"), and *smorzando* ("dying away").

Although is is clear that these annotations are a vital part of the composition, they are not always unequivocal. Their precise interpretation may vary from one composer to the other, which makes it a topic of historical and musicological study. (See Rosenblum [6] for an in depth discussion of the interpretation of dynamics markings in the works of different composers.)

Another relevant question concerns the role of dynamics markings. In some cases, dynamics markings may simply reinforce an interpretation that musicians regard as natural, by their acquaintance with a common performance practice. In other words, some annotated markings may be implied by the structure of the music. In other cases, the composer may annotate highly specific and non-obvious markings, and even fingerings, to ensure the performance achieves the intended effect. An example of this is the music of Beethoven.

The research presented here is intended to help clarify the interpretation of dynamics markings, and how these markings shape the loudness of the performance, in interaction with other aspects of the music. Very generally speaking, the aim is to develop a new methodology for musicological research, that takes advantage of the possibilities of digitized musical corpi, and of advances in statistics and

machine learning – an aim shared with Beran and Mazzola [7].

In a more specific sense, our research follows an intuition that underlies many studies of musical expression, namely that musical expression consists of a number of individual factors that jointly determine what the performance of a musical piece sounds like [8]. The goal is then to identify which factors can account for expression, in casu loudness variations, and to disentangle their contributions to the loudness of the performance.

To this end we present a rather simple model of expressive loudness variations, based on the idea of signal decomposition in terms of basis functions. The primary goal of this approach is to quantify the influence of dynamics markings on the loudness of a performance, but the model is general enough to allow for the inclusion of a wide range of features other than dynamics markings, such as pitch and motivic structure.

The outline of the paper is as follows: In section 2, we describe the model, and the basis functions used in the model. In section 3, we show how the model is used to represent loudness variations in real performances, and perform experiments to evaluate the predictive value of the model, as trained on the data. The results are discussed in section 4, and conclusions and future work can be found in section 5.

2. A LINEAR BASIS MODEL OF EXPRESSIVE DYNAMICS

As stated in the previous section, the model reflects the idea that different aspects of the music jointly shape variation in loudness. When we ignore the possibly complex ways in which aspects might interact, and assume that the eventual loudness is a weighted mixture of these aspects, a linear basis model is an obvious choice to model loudness variation. In this model, one basis function is created for each dynamics annotation, and the performed loudness is regarded as a linear combination of these basis-function. Even if it is simple, we believe that this approach captures the notion that part of the interpretation of a dynamics marking is constant, and that the degree to which the dynamics marking is followed by the performer, is variable. For example, the annotation *p* indicates that the passage spanned by the range of that *p* is to be played softly. This can be represented by a basis-function that removes a constant amount from the loudness curve over that range. The weight of that basis-function determines *how much* the performance gets softer. We will illustrate this further in subsection 2.1.

2.1 Mapping from score to basis

We distinguish between three categories of dynamics annotations, based on their scope, as shown in table 1. The first category, *constant*, represents markings that indicate a particular loudness character for the length of a passage. The passage is ended either by a new *constant* annotation, or the end of the piece. *Impulsive* annotations indicate a change of loudness for only a brief amount of time, usually only the notes over which the sign is annotated. The last

Category	Examples
Constant	*f*, *ff*, *fff*, *mf*, *mp*, *p*, *pp*, *ppp*, *vivo*, *agitato*, *appassionato*, *con anima*, *con forza*, *con fuoco*, *dolce*, *dolcissimo*, *espressivo*, *leggiero*, *leggierissimo*
Impulsive	*fz*, *sf*, *sfz*, *fp*
Gradual	*calando*, *crescendo*, *decrescendo*, *diminuendo*, *smorzando*, *perdendosi*

Table 1. Three categories of dynamics markings

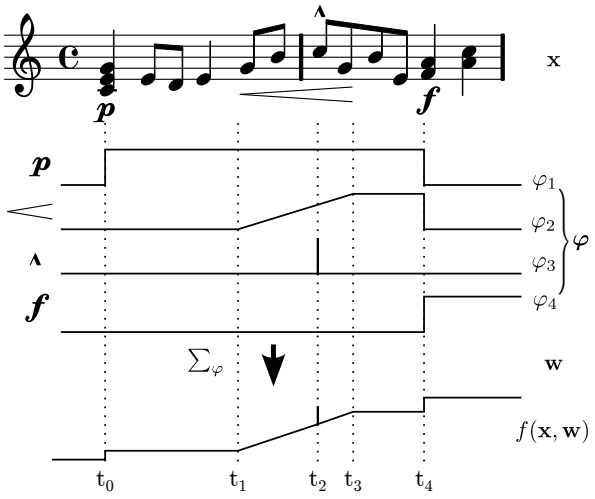

Figure 1. Example of basis functions representing dynamics annotations

category contains those annotations that indicate a gradual change from one loudness level to the other. We call these annotations *gradual*.

Based on their interpretation, as described above, we assign a particular basis function to each category. The constant category is modeled as a step function that has value 1 over the affected passage, and 0 elsewhere. Impulsive annotations are modeled by a unit impulse function, which has value 1 at the time of the annotation and 0 elsewhere. Lastly, gradual annotations are modeled as a combination of a ramp and a step function. It is 0 until the start of the annotation, linearly changes from 0 to 1 between the start and the end of the indicated range of the annotation (e.g. by the width of the 'hairpin' sign indicating a crescendo), and maintains a value of 1 until the time of the next constant annotation, or the end of the piece.

As an illustration, figure 1 shows a fragment of notated music with dynamics markings, and the corresponding basis-functions. The bottom-most curve is a weighted sum of the basis functions φ_1 to φ_4 (using unspecified weights).

2.1.1 Other types of basis-functions

The basis functions shown in figure 1 represent dynamics annotations, and are functions of score time. That is, when two or more notes have the same onset time, the value of

the basis function of these notes will be equal. However, we can also conceive of basis-functions more generally, as functions of the note itself, that may yield different values for notes even if their onset times coincide. This generalization allows us to represent a much larger range of score information as basis-functions.

We will briefly discuss three features that we will include in the model in the form of basis-functions. Firstly, we can include information about the decorative role of notes into the model, by defining a basis function that acts as an indicator function. This function evaluates to 1 for notes that have been marked in the score as grace notes, and to 0 otherwise.

Furthermore, we include a polynomial pitch model into our linear basis model, simply by adding basis functions that map each note to powers of its pitch value. For instance, using the midi note number representation of pitches, the four basis-functions of a third order polynomial pitch model would map a note with pitch 72 to the vector $(72^0, 72^1, 72^2, 72^3) = (1, 72, 5184, 373248)$. There is no need to treat the coefficients of this model separately – by aggregating the polynomial pitch basis functions into the overall model, the coefficients of the polynomial model simply are a subset of the weights of the model. Obviously, the ranges of the polynomial basis-functions will be very diverse. Therefore, in order to keep the model weights in roughly the same range, it is convenient to normalize all basis-functions to the interval $[0, 1]$.

Lastly, we include a more complex feature, based on Narmour's Implication-Realization model of melodic expectation [9]. This model allows for an analysis of melodies that includes an evaluation of the degree of 'closure' occurring at each note [1] . Closure can occur for example due to metrical position, completion of a rhythmic or motivic pattern, or resolution of dissonance into consonance. We use an automatic melody parser that detects metric and rhythmic causes of closure [10]. The output of this parser allows us to define a basis-function that expresses the degree of closure at each note.

Note that we cannot say in advance whether the inclusion of such features will improve our model. By including the features into the model as basis functions we merely create a possibility for the model to explain loudness variations as a function of those features.

2.2 The model

To specify a linear basis model of expressive dynamics, we represent a musical performance as a list of pairs $((x_1, y_1), \cdots, (x_n, y_n))$, where n is the number of notes in the performance, x_i is a representation of the score attributes of the i-th note, and y_i is the loudness value of the i-th note in the performance. We will refer to the vector $(x_1, \cdots, x_n)$ as $\mathbf{x}$, and to the vector $(y_1, \cdots, y_n)$ as $\mathbf{y}$.

We then define a basis function as a function $\varphi_k(.)$ that takes the n elements of $\mathbf{x}$ as arguments to produce a real valued vector of size n. Once a set $\boldsymbol{\varphi} = (\varphi_1(.), \cdots, \varphi_m(.))$ of m basis functions is fixed, it can be applied to a musical score $\mathbf{x}$ to yield a matrix $\boldsymbol{\varphi}(\mathbf{x}) = (\varphi_1(\mathbf{x}), \cdots, \varphi_m(\mathbf{x}))$ of size $n \times m$, where n is the number of notes in $\mathbf{x}$.

The definition of the elements of $\mathbf{x}$ is not mathematically relevant, since $\mathbf{x}$ will only appear as an argument to $\boldsymbol{\varphi}$. Suffice it to say that the elements of $\mathbf{x}$ contain basic note information such as notated pitch, onset time and offset time, and any dynamics markings that are annotated in the score. [2]

The model is defined as a function y of the score $\mathbf{x} = (x_1, \cdots, x_n)$ and a vector of weights $\mathbf{w} = (w_1, \cdots, w_m)$, such that the loudness is a linear combination of the basis functions:

$$f(\mathbf{x}, \mathbf{w}) = \mathbf{w}^T \boldsymbol{\varphi}(\mathbf{x}) \tag{1}$$

Thus, for note x_i, the predicted loudness is computed as:

$$\hat{y_i}(\mathbf{w}) = f(x_i, \mathbf{w}) = \mathbf{w}^T \boldsymbol{\varphi}(x_i) = \sum_j^m w_j \varphi_j(x_i) \tag{2}$$

2.3 Learning and prediction with the linear basis model

Given performances in form $(\mathbf{x}, \mathbf{y})$ we can use the model in equation (1) to estimate the weights $\mathbf{w}$, which is a simple linear regression problem. The most common approach to this kind of problem is to compute $\mathbf{w}$ as the least squares solution[11], that is, the $\mathbf{w}$ that minimizes the sum of the squared differences between the loudness predictions $y(\mathbf{x}, \mathbf{w})$ of the model and the observed loudness $\mathbf{y}$:

$$\mathbf{w}_{\mathbf{x},\mathbf{y}} = \text{argmin}_{\mathbf{w}} \sum_i^n [\, y_i - \hat{y}_i(\mathbf{w}) \,]^2 \tag{3}$$

To find the optimal $\mathbf{w}$ for a set of musical performances $((\mathbf{x}_1, \mathbf{y}_1), \cdots, (\mathbf{x}_K, \mathbf{y}_K))$, we can use two different approaches. One way is to compute a vector $\mathbf{w}_{\mathbf{x}_k,\mathbf{y}_k}$ for each performance k according to equation (3), and combine the $\mathbf{w}_{\mathbf{x}_k,\mathbf{y}_k}$'s in some way to form a final estimate of $\mathbf{w}_{\mathbf{x},\mathbf{y}}$, e.g. by taking the median for each weight w_j

Another approach is to concatenate the respective $\mathbf{x}_k$'s and $\mathbf{y}_k$'s of the performances into a single pair $(\mathbf{x}, \mathbf{y})$, and to find $\mathbf{w}_{\mathbf{x},\mathbf{y}}$ directly according to equation (3). For this to work however, $\boldsymbol{\varphi}$ must have the same number of columns, i.e. the same number of basis functions, for each $\mathbf{x}_k$. This may or may not be the case, depending on how we define our basis functions. In the following paragraph, we briefly explain two alternative approaches to defining basis functions.

2.3.1 Local and global bases

The mapping of dynamics annotations can be done in different ways. The first possibility is to define a new basis function for each dynamics marking that we encounter in the score. This means for example, that repeated *crescendo* annotations are represented in different basis functions, and

[1] Narmour's concept of closure is subtly different from the common notion of musical closure in the sense that the latter refers to 'ending' whereas the former refers to the inhibition of the listener's expectation of how the melody will continue. In spite of the difference in meaning, both notions are arguably related.

[2] the representation of dynamics markings can be realized for example by indicator functions over the elements of $\mathbf{x}$

that each *crescendo* can be fitted to the observed loudness independent of the other *crescendi*. This approach leads to basis functions that are zero throughout the piece, except for the passage where the corresponding annotation applies. We call such basis functions *local*. They have the benefit that they make the model more flexible, and therefore allow for better approximations of the data. A drawback of this method is that we obtain as many weight parameters as we have *crescendi*, rather than a single weight estimation for the *crescendo* sign in general. Also, as the number of basis functions we obtain for a musical piece varies, depending on the number of dynamics annotations in the score, the fitting method by concatenating musical performances, as proposed above, becomes impossible.

Alternatively, we can choose to assign a single basis function to each type of marking. This implies for example, that we combine the different step functions of each *p* in the piece into a single function, e.g. by summing them up. We call this a *global* basis function.

Note that the basis functions for the other features mentioned (features related to pitch, grace notes, and I-R closure) are all global: for each feature the values of all notes in the performance are aggregated into a single basis function. There is therefore a fixed set of basis-functions representing the global features, independent of the piece.

2.3.2 Prediction with local and global bases

In the global case, once a weight vector $\mathbf{w}$ has been learned from a data set D, predictions for a new piece $\mathbf{x}$ can be made easily, by computing the matrix $\boldsymbol{\varphi}(\mathbf{x})$ and subsequently the dot product $f(\mathbf{x}, \mathbf{w}) = \mathbf{w}^T \boldsymbol{\varphi}(\mathbf{x})$.

In the case where dynamics annotations are modeled by local basis functions, the $\mathbf{w}$'s that have been learned for each piece in the training set may have different lengths. In this case, we split up the $\mathbf{w}$'s of each training piece into a set of weights $\mathbf{w}_A$ that correspond to global basis functions (i.e. for the non-dynamics features), and the remaining weights $\mathbf{w}_B$, for the dynamics annotations, which vary in number. Over the weights $\mathbf{w}_A$ (which have fixed size), we take the median. The weights $\mathbf{w}_B$ of all pieces in the training set are pooled. This pool includes multiple weights for each dynamics marking. To predict weights for the dynamics markings of a test piece, we use a support vector machine [12], that has been trained on the pool of weights $\mathbf{w}_B$ from the training data. The medians over the $\mathbf{w}_A$'s and the SVM predictions from the pool of $\mathbf{w}_B$'s are then appended to yield the final vector $\mathbf{w}$ used for predicting the loudness of the new piece.

3. MODEL EVALUATION

To evaluate the different features, and basis modeling approaches we have discussed above, we use it to model and predict the loudness in a set of real performances. Details of the data set are given in subsection 3.1. We wish to highlight that this data set is larger than any other data set we know of, that has a similar level of recording precision (exact onset times and loudness for each performed note).

We evaluate two aspects of the model variations. The first is the goodness-of-fit, that is, how well can the model represent the data (subsection 3.2). The second is the predictive accuracy (subsection 3.3). In both cases, we compare different features sets, for both basis modeling approaches we discussed, *local* and *global*.

We use the following abbreviations to refer to the different kinds of features: DYN: dynamics annotations. These annotations are represented by one basis function for each marking in table 1, plus one basis function for accented notes; PIT: a third order polynomial pitch model (3 basis functions) [3] ; GR: the grace note indicator basis; IR: two basis-functions, one indicating the degree of closure, and another representing the squared distance from the nearest position where closure occurs. The latter feature forms arch-like parabolic structures reminiscent of Todd's model of dynamics [3].

The total number of parameters in the model is thus 30 (DYN) + 3 (PIT) + 1 (GR) + 2 (IR) + 1 (constant basis) = 37, or less, depending on the subset of features that we choose. In the evaluation, we omit the feature combinations that consist of only GR and IR, since we expect their influence on loudness to be marginal with respect to the features DYN and PIT.

3.1 Data Set

For the evaluation we use the Magaloff corpus [13] – a data set that comprises live performances of virtually the complete Chopin piano works, as played by the Russian-Georgian pianist Nikita Magaloff (1912-1992). The music was performed in a series of concerts in Vienna, Austria, in 1989, on a Bösendorfer SE computer-controlled grand piano [14] that recorded the performances onto a computer hard disk. The data set comprises more than 150 pieces, adding up to almost 10 hours of music, and containing over 330,000 performed notes. These data, which are stored in a native format by Bösendorfer, were converted into standard MIDI format, representing loudness values as a parameter named *velocity*, taking values between 0 (silent), and 127 (loudest). For the purpose of this experiment, velocity values have been transformed to have zero-mean per piece.

Information about dynamics markings in the score was obtained from optical music recognition from the scanned musical scores (see [13] for details). We have used the Henle Urtext Edition wherever possible.

3.2 Goodness-of-fit of the loudness representation

To quantify how well the model is able to capture loudness variations of the performances. We compute the optimal weight vector $\mathbf{w}$ of the model for each piece in the data set. In the global case, a single weight vector is computed on the whole data set, and is applied to the basis $\boldsymbol{\varphi}(\mathbf{x})$ of each piece $\mathbf{x}$. In the local case, a $\mathbf{w}$ was computed for each piece, and used to to fit the model to the data. This is done for the different features and combinations discussed at the beginning of section 3.

[3] The chosen polynomial order of 3 was chosen as most appropriate, after a visual inspection of scatterplots showing the relationship between loudness and pitch. The constant basis function that is part of the polynomial pitch model is omitted because it is subsumed by a default constant basis function included in every basis combination.

	r		R^2	
Basis (global)	avg.	std.	avg.	std.
DYN	0.332	(0.150)	0.133	(0.117)
PIT	0.456	(0.108)	0.219	(0.097)
DYN+PIT	0.565	(0.106)	0.330	(0.122)
DYN+PIT+GR	0.567	(0.107)	0.332	(0.123)
DYN+PIT+IR	0.575	(0.102)	0.341	(0.120)
DYN+PIT+GR+IR	0.577	(0.102)	0.343	(0.120)
Basis (local)				
DYN	0.497	(0.170)	0.276	(0.160)
PIT	0.456	(0.108)	0.219	(0.097)
DYN+PIT	0.670	(0.113)	0.462	(0.146)
DYN+PIT+GR	0.671	(0.113)	0.463	(0.146)
DYN+PIT+IR	**0.678**	(0.109)	0.471	(0.142)
DYN+PIT+IR+GR	**0.678**	(0.109)	**0.472**	(0.142)

Table 2. Goodness of fit of the model; See section 3 for abbreviations

	r		R^2	
Basis (global)	avg.	std.	avg.	std.
DYN	0.192	(0.173)	0.020	(0.100)
PIT	0.422	(0.129)	0.147	(0.111)
DYN+PIT	**0.462**	(0.125)	0.161	(0.156)
DYN+PIT+GR	**0.462**	(0.125)	0.161	(0.156)
DYN+PIT+IR	**0.462**	(0.124)	0.162	(0.155)
DYN+PIT+GR+IR	**0.462**	(0.124)	0.162	(0.154)
Basis (local)				
DYN	0.192	(0.179)	0.024	(0.109)
PIT	0.415	(0.137)	0.149	(0.149)
DYN+PIT	0.459	(0.126)	0.151	(0.220)
DYN+PIT+GR	0.459	(0.123)	0.153	(0.195)
DYN+PIT+IR	0.455	(0.130)	0.141	(0.231)
DYN+PIT+IR+GR	0.457	(0.123)	**0.188**	(0.126)

Table 3. Predictive accuracy the model in a leave-one-out scenario; See section 3 for abbreviations

The results of this evaluation are shown in table 2. The goodness-of-fit is expressed in two quantities: r is the Pearson product-moment correlation coefficient, denoting how strong the observed loudness, and the loudness values of the fitted model correlate. The quantity R^2 is the coefficient of determination, which is defined as:

$$R^2 = 1 - \frac{SS_{err}}{SS_{obs}}, \tag{4}$$

where:

$$SS_{obs} = \sum_i^n (y_i - \bar{y})^2, \quad SS_{err} = \sum_i^n (y_i - \hat{y}_i(\mathbf{w}))^2. \tag{5}$$

The coefficient of determination is a measure for how much of the loudness variance is accounted for by the model. In the case of a perfect fit $R^2 = 1$, since $SS_{err} = 0$. In the undesirable case where the variance of the loudness increases by subtracting the model fit from the observations, we have $SS_{err} > SS_{obs}$, and R^2 will be negative. Table 2 lists the average and standard deviations of the r and R^2 values over 154 musical pieces.

The results show that both the strongest correlation, and the highest coefficient of determination is achieved when using local basis for dynamics markings, and including all features. This is unsurprising, since in the global setting a single weight vector is used to fit all pieces, whereas in the local setting each piece has its own weight vector. Furthermore, since adding features increases the number of parameters in the model, it will also increase the goodness-of-fit.

3.3 Predictive accuracy of the model

The additional flexibility of the model, by using local bases and adding features, may increase its goodness-of-fit. However, it is doubtful that it will help to obtain good model predictions for unseen musical pieces. To evaluate the accuracy of the predictions of a trained model for an unseen piece, we perform a leave-one-out cross-validation over the 154 pieces. The predictions are evaluated again in terms of averaged r and R^2 values over the pieces, which are shown in table 3.

The average correlation coefficients between prediction and observation for the local and global basis settings are roughly similar, ranging from weak ($r = .19$) to medium correlation ($r = .46$). In the global setting, increasing the complexity of the model does not affect its predictive accuracy, whereas in the local setting, maximal predictive accuracy is achieved for models of moderate complexity (including dynamics, pitch, and grace note information). The decrease of accuracy for more complex models is likely to be caused by overfitting.

Interestingly, the highest proportion of explained variance ($R^2 = .19$) is achieved by the predictions of the local model with all available features (DYN+PIT+IR+GR). However, it should be noted that the standard deviation of R^2 is rather large in most cases, indicating that for some pieces a much larger proportion of the loudness variance can be explained than for others.

4. DISCUSSION OF RESULTS

The results presented in the previous section show a substantial difference in the contribution of dynamical annotations (DYN) and pitch (PIT) to the performance of the model. The fact that pitch explains a larger proportion of the loudness variance than the dynamics annotations may come as a surprise, given that dynamics annotations are by nature intended to guide variations in loudness.

Although the data set spans a large set of performances, it is important to realize that the results are derived from performances of a single performer, performing the music of a single composer. The importance of pitch as a predictor for loudness may be different for other performers, composers, and musical genres. Specifically, we hypothesize that the fact that pitch has a strong predictive value for loudness in our data set may be a consequence of *melody*

lead. This phenomenon, which has been the subject of extensive study (see [15, 16]), consists in the consistent tendency of pianists to play melody notes both louder and slightly earlier than the accompaniment. This makes the melody more clearly recognizable by the listener, and may improve the sensation of a coherent musical structure. In many musical genres (though not all), the main melody of the music is expressed in the highest voice, which explains the relationship between pitch loudness.

This effect is clearly visible in figure 2, which displays observed, fitted, and predicted loudness for the final measures of Chopin's Prelude in B major (Opus 28, Nr. 11). In this plot, the loudness of simultaneous notes is plotted at different (adjacent) positions on the horizontal axis, for the ease of interpretation. Melody notes are indicated with dotted vertical lines. It is easily verified by eye that the loudness of melody notes is substantially higher than the loudness of non-melody notes. This effect is very prominent in the predictions of the model as well. [4]

Although observed and predicted loudness are visibly correlated, figure 2 shows that the variance of the prediction is substantially lower than that of the observation, meaning that expressive effects in the predicted performance are less pronounced. The lower variance is most likely caused by the fact that the (relatively small set of) model parameters has been optimized to performances of a wide range of different pieces, preventing the model from accurately capture loudness variance for individual performances. This problem may require a more sophisticated model, or alternatively, a separate treatment of musical pieces with distinct musical characters.

In spite of this, the results generally show that using a simple linear basis model, it is possible to capture a substantial proportion of loudness variations, both in function of dynamics annotations in the score and as a consequence of more implicit phenomena such as melody lead. We believe that this kind of model can provide a general methodology to study the factors that influence musical expression.

The model may also be of use to model variation in the articulation of notes. However, the applicability of the model to other aspects of musical expression, may not be straight-forward in all cases. For example, expressive timing (e.g. in terms of inter-onset interval (IOI) ratios) is a phenomenon that affects the time dimension of the performance. Therefore, it is not desirable to predict IOI values independently for simultaneous notes.

5. CONCLUSIONS AND FUTURE WORK

The work presented in this paper corroborates a growing insight in music performance research: that even if musical expression is a highly complex and subjective phenomenon, it is by no means fully unsystematic. We have shown that using a simple linear basis model, we can generate loudness predictions from musical scores that show substantial positive correlation with loudness as observed in human performances by a professional pianist.

The model has several advantages. Firstly, it embodies the common intuition that expression in music performance is a result of multiple factors that jointly determine how the performance sounds. Secondly, it is concise: with 37 parameters, it is possible to explain almost 29% of the loudness variance in a data set of over 330,000 performed notes (table 2).

Improvements to the model can be conceived at different fronts. For example, a more sophisticated approach may be taken to infer the weight vectors from a data set. In particular, a Bayesian approach seems attractive, in which a prior probability distribution over weights is specified. Another improvement would be to learn basis-functions from the data, or adapt manually specified basis-functions. For this, techniques developed in the field of dictionary learning, such as *matching pursuit*, might be used. Finally, it is desirable to assess the quality of predicted loudness curves by subjective evaluation through listening tests, in addition to numerical comparison of predictions with target performances.

[4] Sound examples of musical fragments with loudness predicted by the model can be found at `www.cp.jku.at/research/TRP109-N23/BasisMixer/midis.html`

Acknowledgments

This research is supported by the Austrian Research Fund (FWF, Z159 "Wittgenstein Award"). We are indebted to Mme. Irene Magaloff for her generous permission to use her late husbands performance data for our research. We are grateful to Sebastian Flossmann for his effort in the preparation of the Magaloff corpus. For this research, we have made extensive use of free software, in particular R, python, and GNU/Linux.

6. REFERENCES

[1] E. F. Clarke, "Generative principles in music," in *Generative Processes in Music: The Psychology of Performance, Improvisation, and Composition*, J. Sloboda, Ed. Oxford University Press, 1988.

[2] C. Palmer, "Music performance," *Annual Review of Psychology*, vol. 48, pp. 115–138, 1997.

[3] N. Todd, "The dynamics of dynamics: A model of musical expression," *Journal of the Acoustical Society of America*, vol. 91, pp. 3540–3550, 1992.

[4] R. Parncutt, *Perspektiven und Methoden einer Systemischen Musikwissenschaft*. Germany: Peter Lang, 2003, ch. Accents and expression in piano performance, pp. 163–185.

[5] B. H. Repp, "Diversity and commonality in music performance - An analysis of timing microstructure in Schumann's "Träumerei"," *Journal of the Acoustical Society of America*, vol. 92, no. 5, pp. 2546–2568, 1992.

[6] S. P. Rosenblum, *Performance practices in classic piano music: their principles and applications*. Indiana University Press, 1988.

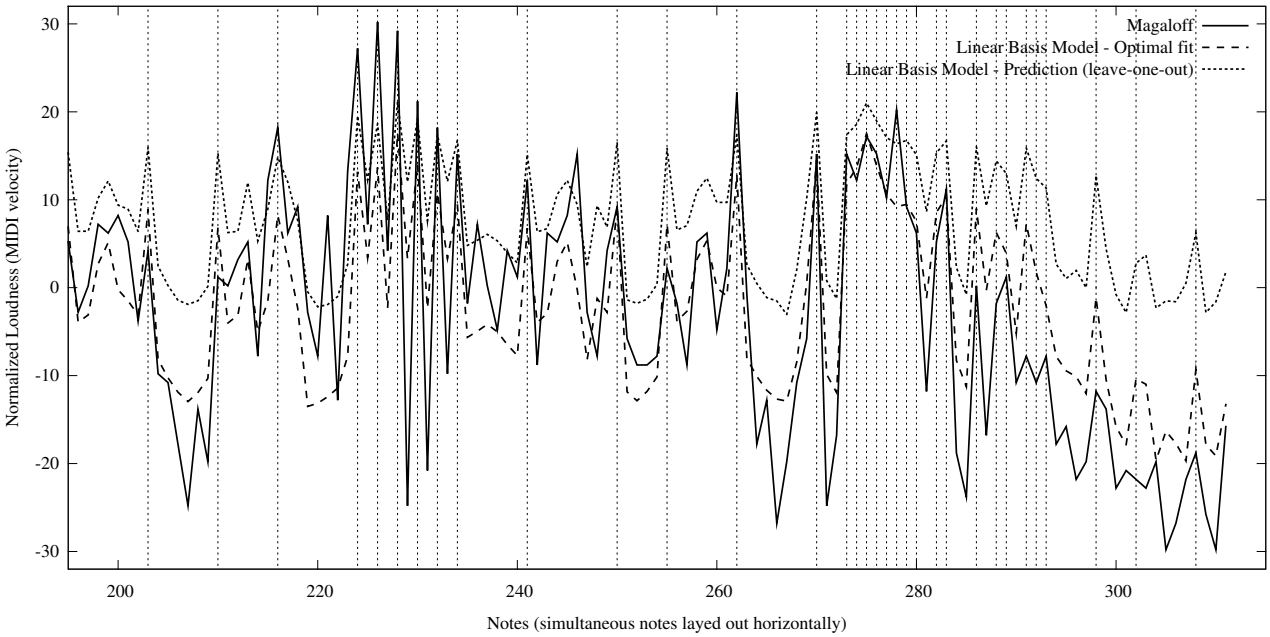

Figure 2. Observed, fitted, and predicted note-by-note loudness of Chopin's Prelude in B major (Opus 28, Nr. 11), from measure 16 onwards; Fitting and prediction was done using the global basis DYN+PIT+GR+IR (see section 3); Vertical dotted lines indicate melody notes

[7] J. Beran and G. Mazzola, "Analyzing musical structure and performance— a statistical approach," *Statistical Science*, vol. 14, no. 1, pp. 47–79, 1999.

[8] C. Palmer, "Anatomy of a performance: Sources of musical expression," *Music Perception*, vol. 13, no. 3, pp. 433–453, 1996.

[9] E. Narmour, *The analysis and cognition of basic melodic structures : the Implication-Realization model*. University of Chicago Press, 1990.

[10] M. Grachten, "Expressivity-aware tempo transformations of music performances using case based reasoning," Ph.D. dissertation, Pompeu Fabra University, Barcelona, Spain, 2006, ISBN: 635-07-094-0.

[11] A. Björck, *Numerical methods for least squares problems*. SIAM, 1996.

[12] B. Boser, I. Guyon, and V. Vapnik, "A training algorithm for optimal margin classifiers," in *Fifth Annual Workshop on Computational Learning Theory*. Pittsburgh: ACM, 1992, pp. 144–152.

[13] S. Flossmann, W. Goebl, M. Grachten, B. Niedermayer, and G. Widmer, "The Magaloff Project: An Interim Report," *Journal of New Music Research*, vol. 39, no. 4, pp. 369–377, 2010.

[14] R. A. Moog and T. L. Rhea, "Evolution of the Keyboard Interface: The Bösendorfer 290 SE Recording Piano and the Moog Multiply-Touch-Sensitive Keyboards," *Computer Music Journal*, vol. 14, no. 2, pp. 52–60, 1990.

[15] B. Repp, "Patterns of note onset asynchronies in expressive piano performance," *Journal of the Acoustical Society of America*, vol. 100, no. 6, pp. 3917–3932, 1996.

[16] W. Goebl, "Melody lead in piano performance: expressive device or artifact?" *Journal of the Acoustical Society of America*, vol. 110, no. 1, pp. 563–572, 2001.