Linear basis models for prediction and analysis of musical expression

Maarten Grachten[1]
Gerhard Widmer[1,2]

[1] Department of Computational Perception
Johannes Kepler University, Linz, Austria

[1,2] 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. Several models exist that explain expressive variations over the course of a performance, for example in terms of phrase structure, or musical accent. Often however expressive markings are written explicitly in the score to guide performers. We present a modelling framework for musical expression that is especially suited to model the influence of such markings, along with any other information from the musical score. In two separate experiments, we demonstrate the modelling framework for both predictive and explanatory modelling. Together with the results of these experiments, we discuss our perspective on computational modelling of musical expression in relation to musical creativity.

1 Introduction and related work

When a musician performs a piece of notated music, the performed music typically shows large variations in expressive parameters like tempo, dynamics, articulation, and depending on the nature of the instrument, further 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 affective elements, and also elements that convey musical structure (Clarke, 1988; Palmer, 1997).

These insights have led to numerous models of musical expression. The aim of these models is to explain the variations in expressive parameters as a function of the performer's interpretation of the music, and most of them can roughly be classified as either focusing on affective aspects of the interpretation, or structural aspects. An example of the former is the model of Canazza et al. (2004, 2002), which associates perceptual dimensions of performance to physical sound attributes. They identify sensorial and affective descriptions of performances along these dimensions. Furthermore, the rule based model of musical expression used by Bresin and Friberg (2000) allows for modelling both structure and affect related expression. For the latter, they use the notion of *rule palettes* to model different emotional interpretations of music. These palettes determine the strength of each of a set of predefined rules on how to perform the music.

A clearly structure-oriented approach is the model by Todd (1992), in which tempo and dynamics are (arch-shaped) functions of the phrase structure of the piece. Another example is Parncutt's (2003) model of musical accent, which states that expression is a function of the musical salience of the constituents of a piece. Timmers et al. (2002) propose a model for the timing of grace notes. Lastly, Tobudic and Widmer (2003) use a combination of case based reasoning and inductive logic programming to predict dynamics and tempo of classical piano performances, based on a phrase analysis of the piece, and local rules learnt from data.

A structural aspect of the music that has been remarkably absent in models of musical expression, are expressive markings written in the score. Many musical scores, especially those by composers from the early romantic era, include instructions for the interpretation of the notated music. Common instructions concerning the dynamics of the performance include *forte* (*f*) and *piano* (*p*), indicating loud and soft passages, respectively, and *crescendo/decrescendo* for a gradual increase and decrease in loudness, respectively. Some less common markings prescribe a dynamic evolution in the form of a metaphor, such as *calando* ("growing silent"). These metaphoric markings may pertain to variations in one or more expressive parameters simultaneously.

At first sight, the lack of expressive markings as a (partial) basis for modelling musical expression might be explained by the fact that, since the markings appear to prescribe the expressive interpretation explicitly, modelling is trivial, and therefore without scientific value. However, modelling the influence of expressive markings is far from trivial, for various reasons. Firstly, expressive markings are not always unequivocal. Their interpretation may vary from one composer to the other, which makes it a topic of historical and musicological study (Rosenblum, 1988). 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. That is, some annotated markings may be implied by the structure of the music. In other cases, the composer may annotate markings precisely at non-obvious places.

When dealing with historical recordings in empirical studies of musical interpretation, a practical difficulty with expressive markings is that in many cases, it is unknown which edition of the score (if any) the performer used. Often several editions of musical scores exist, and these editions may have different expressive annotations, due to revisions by the composer, music educators, or the publisher.

Another challenging fact is that even when the corresponding edition of the score for a performance is known, it lies within the artistic freedom of the performer to interpret annotations differently, play them with modifications, or ignore them altogether. Even if this complicates a straight-forward approach to modelling the effect of expressive annotations, this freedom forms part of the basis for music performance as a creative activity. In that sense, a model that captures *how* a musician deals with performance annotations in the score, can be regarded as a description of the musician's creative behaviour. Such a model however is not a model of the creative process itself, but of an artifact resulting from a creative process.

In this paper, we describe a framework that allows for modelling, among other things, the effect of annotated expressive markings on music performances. This framework 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 (Palmer, 1996). With this framework, expressive information from human performances can be decomposed into (for now predefined) components, by fitting the parameters of a linear model to those performances. We will refer to this as the linear basis modelling (LBM) framework.[1]

Learnt models serve both predictive and explanatory purposes. As a predictive tool, models find practical application in tasks such as automatic score-following and accompaniment. In an explanatory setting, a model fitted to data reveals how much of the variance in an expressive parameter is explained by each of the basis functions (representing structural aspects of the musical score).

The outline of the paper is as follows: In section 2, we describe the LBM framework, and discuss possible types of basis functions. The experimentation (section 3) consist of two parts: In subsection 3.1, we show how the model is used to represent dynamics in real performances, and perform experiments to evaluate the predictive value of the model, as trained on the data. In subsection 3.2, we use fitted models to quantitatively assess differences between the way pianists interpret expressive markings. The results are presented and discussed in section 4. In that section, we also relate our approach to the question of creativity in the context of musical expression. Conclusions and future work are presented in section 5.

2 Linear basis models of musical expression

As stated in the introduction, a common view is that variation in the expressive parameters of music is shaped jointly by a variety of different structural and affective aspects of the music, in combination with the performer's expressive intentions. Depending on these intentions, such aspects may determine expres-

[1] We use the term *framework* to refer to the general modelling methodology, including techniques to estimate parameters, and to predict new performances. By *model*, we mean an instantiation of this methodology, using a fixed selection of basis-functions.

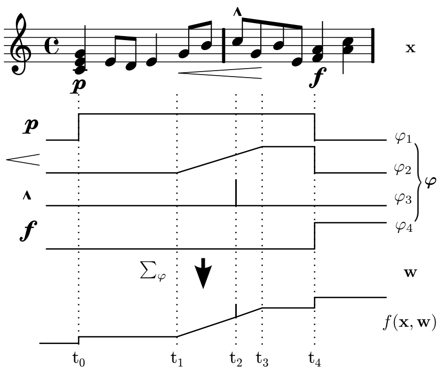


Figure 1: Example of basis functions representing dynamics annotations

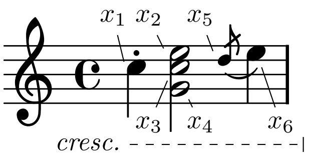
sive variations directly, but it is also likely that they shape the performance through highly complex interactions.

The purpose of the LBM framework is to explore the simpler relationships between these aspects of the music and its expressive performance. It does so by relying on several strongly simplifying assumptions. Firstly, each expressive parameter depends only on score information; this implies that both mutual dependencies between expressive parameters and temporal dependencies within parameters are not modelled explicitly. Secondly, as the name suggests, expressive parameters are modelled as depending linearly on score information.

Before we describe the notion of basis functions representing score features, and the LBM framework, we provide an illustration in figure 1, to clarify the general idea. The figure shows a fragment of notated music with dynamics markings. The first four curves below the notated music represent each of the four markings as basis functions. The basis functions evaluate to zero where the curves are at their lowest, and to one where they are at their highest. Note that each of the basis functions is only non-zero over a limited range of time, namely the time where the corresponding dynamic marking takes effect in the music. The bottom-most curve is a weighted sum of the basis functions φ_1 to φ_4 (using unspecified weights $\mathbf{w}$), that represents an expressive parameter, in this case the note dynamics.

2.1 Representation of score information as basis functions

In the past, the *MIDI* format has been often used as a representation scheme for musical scores. Being intended as a real-time communication protocol however, this scheme is not very suitable for describing structural score information beyond the pitches and onsets of notes. By now, the more descriptive *MusicXML* format (Good, 2001) is widely used for distributing music. The types of information we will refer to in this subsection are all contained in a typical MusicXML representation of a musical piece.



note	$\varphi_{staccato}(.)$	$\varphi_{grace}(.)$	$\varphi_{pitch}(.)$	$\varphi_{pitch^2}(.)$	$\varphi_{crescendo}(.)$
x_1	1	0	$\frac{72}{127}$	$(\frac{72}{127})^2$	0.00
x_2	0	0	$\frac{76}{127}$	$(\frac{76}{127})^2$	0.25
x_3	0	0	$\frac{72}{127}$	$(\frac{72}{127})^2$	0.25
x_4	0	0	$\frac{67}{127}$	$(\frac{67}{127})^2$	0.25
x_5	0	1	$\frac{74}{127}$	$(\frac{74}{127})^2$	0.75
x_6	0	0	$\frac{76}{127}$	$(\frac{76}{127})^2$	0.75

Figure 2: A score fragment illustrating various kinds of basis functions (see text for explanation)

We define a musical score as a sequence of elements that hold information, and may also refer to other elements. For our purposes, it is relevant to distinguish between note elements, and non-note elements. Note elements hold local information about individual notes, such as pitch, onset time, and duration information, but also any further annotations that describe the note, such as whether the note has a staccato sign, an accent, a fermata, and whether it is a grace note. Non-note elements represent score information that is not local to a specific note. Examples are expressive markings for dynamics (*p*, *f*, *crescendo*, et cetera), tempo (*lento*, *ritardando*, *presto*, et cetera), but possibly also time and key signatures, and slurs.

The purpose of basis functions for modelling expression is to capture some structural aspect of the score, and express the relation of each score note to that aspect, as a real number between 0 and 1. If we denote the set of all note-elements by $\mathcal{X}$, then a basis function has the form: $\varphi : \mathcal{X} \to [0, 1]$. Although this suggests that the evaluation of a basis function on a note element only depends on that element, many interesting types of basis function take into account the context of the note. Therefore, it is convenient to think of a note element as holding a reference to its context in the piece it occurs in (for example, to determine whether it occurs inside the scope of a *crescendo* sign).

Note that defining basis functions as functions of notes, rather than functions of score time, increases the modelling power considerably. It allows for modelling several forms of musical expression related to simultaneity of musical events. Examples are: the micro-timing of note onsets in a chord (*chord spread*), an expressive device that has hardly been studied empirically; and the accentuation of the melody voice with respect to accompanying voices by playing it louder, and slightly earlier (*melody lead*) (Goebl, 2001).

In the following, we will propose several types of basis functions.

2.1.1 Indicator basis functions for note attributes

The simplest type of basis function is an indicator function that evaluates to one wherever a specific characteristic occurs, and to zero otherwise. For example, we can define a function $\varphi_{staccato}$ for notes that have *staccato* annotations, and a function φ_{grace} for grace notes. Both types of functions are illustrated in figure 2. By including φ_{grace} as a basis function for dynamics, it can account for any systematic deviations in dynamics of performed grace notes. Similarly, $\varphi_{staccato}$ can reasonably be expected to account for part of the variance in note articulation.

2.1.2 Basis function representation of a polynomial pitch model

The motivation to include pitch as a factor for modelling expressive dynamics comes from the observation that in at least two large corpora of piano performances (of different music, and by different performers) there is a statistical dependence of note dynamics, measured as MIDI velocity, on note pitch. Both corpora comprise exact measurements of the dynamics of played notes, through the use of Bösendorfer's computer controlled grand piano (see subsection 3.1.2). [2]

Figure 3 shows the relation between dynamics and the pitch in a scatter plot for two performance corpora. Dynamics and pitch are clearly not statistically independent. Note however, that the relation does not appear to be perfectly linear. Notes with lower pitches on average appear to be played louder than expected based on a linear relationship. To find a good representation for the dynamics-pitch relationship, we have fitted polynomials of different orders to the data (see figure 3). The third order model was selected for its tendency to map higher pitches to relatively moderate velocities, particularly for the Chopin data. [3]

This representation, which we call a *polynomial pitch model*, can be integrated elegantly in the LBM framework, allowing joint estimation of the parameters of the pitch model and the parameters of other basis functions. The inclusion of the pitch model is achieved simply by defining basis functions φ_{pitch}, φ_{pitch^2}, et cetera , that map notes to the respective powers of their (normalised) MIDI pitch numbers. The first and second degree basis functions for pitch are illustrated for the example fragment in figure 2.

2.1.3 Basis functions for expressive markings of dynamics

We distinguish between three categories of dynamics annotations (shown in table 1), based on their meaning. The first category, *constant*, represents markings

[2] The loudness of a note depends on several factors, and the relation between the MIDI velocity of a note performed on the Bösendorfer piano and its loudness is far from straightforward. The relation between sound pressure level and MIDI velocity on computer controlled pianos has been investigated by Goebl and Bresin (2003). For the Bösendorfer piano this relation is roughly linear from MIDI velocities 40 upwards, although it depends on pitch.

[3] Listening to synthesized model predictions revealed that a second order pitch model tends to overemphasize higher pitches.

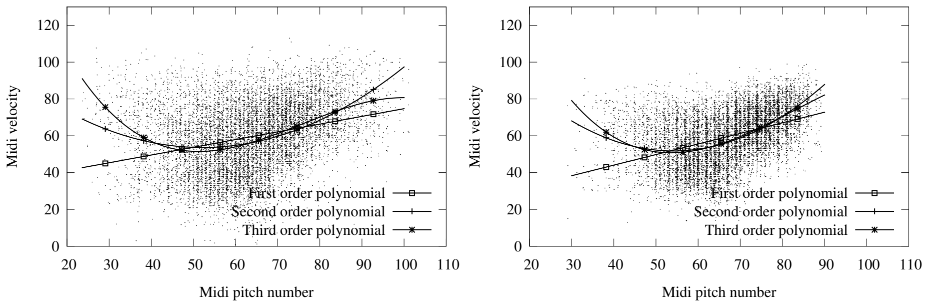


Figure 3: Dependency of dynamics and pitch in Magaloff's performances of Chopin's piano works (left); and Batik's performances of Mozart's piano sonatas (right); Although the displayed lines were fitted on complete data sets; only a subset of the data points are plotted, for convenience

Category	Examples	Basis function
Constant	*f*, *ff*, *p*, *dolce*, *agitato*	step
Impulsive	*fz*, *fp*	impulse
Gradual	*crescendo*, *diminuendo*, *perdendosi*	ramp + step

Table 1: Three categories of dynamics markings

that indicate a particular dynamic 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 sound level for only a brief amount of time, usually only the notes over which the sign is annotated. The last category contains those annotations that indicate a gradual change from one sound 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 modelled as a step function that has value 1 over the affected passage, and 0 elsewhere. Impulsive annotations are modelled by a unit impulse function, which has value 1 for notes at the time of the annotation and 0 elsewhere. Lastly, gradual annotations are modelled 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. Three types of basis functions are illustrated in figure 1, with a more detailed example of the $\varphi_{crescendo}$ function in figure 2.

2.1.4 Implication-Realization based basis-functions

Lastly, we include a more complex feature, based on Narmour's (1990) Implication-Realization model of melodic expectation. This model allows for an analysis of melodies that includes an evaluation of the degree of 'closure' occurring at each note[4]. 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 (Grachten, 2006), and represent the degree of closure at each note in a basis-function.

2.1.5 Global and local bases

An important decision in the design of basis functions is whether a basis function corresponds to a feature in general, or to a single *instance* of that feature. In the case of a grace note basis function for example, the first approach results in a single function that evaluates to one for all grace notes. We call such basis functions *global*. In contrast, the second approach results in one basis-function for every grace-note, evaluating to one *only* on the corresponding grace note. We refer to these as *local* basis functions.

Whether one approach is to be preferred over the other may depend, among other things, on the type of feature, on the purpose of the model, and the amount of data available for fitting the model. Nevertheless, the choice has some general implications. First of all, a local basis modelling approach will lead to more basis functions, and thus to models with more parameters. This provides more flexibility, and will result in better approximations of the expressive target to be modelled. But the larger number of parameters also makes models more prone to overfitting. Apart from that, using local basis functions in general leads to the situation that different pieces are represented by a different number of basis functions, depending on the (number of) annotations present in the score. This makes prediction slightly more complicated (we deal with this in subsection 2.3.1). Nevertheless, it makes sense in some cases to choose local basis functions, for example when the interpretation of features is expected to include outliers, or vary strongly from one instance to the other.

2.2 Model description

As mentioned in the introduction, each of the expressive parameters ($\mathbf{y}$) is modelled separately. LBM is independent of the interpretation of $\mathbf{y}$. For example, in (Grachten and Widmer, 2011) it is used to model dynamics, whereas in (Krebs and Grachten, 2012), $\mathbf{y}$ represents expressive tempo. In this subsection, we will use $\mathbf{y}$ without a specific interpretation, and we will refer to it as the *target*.

The central idea behind LBM is that it provides a way to determine the optimal influence of each of a set of basis functions, in the approximation of

[4] 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.

the target. The influence of a basis-function is expressed by a weight w. This leads to the following formalisation: given a musical score, represented as a list of N notes $\mathbf{x} = (x_1, \cdots, x_N)$, and a set of K predefined basis functions $\boldsymbol{\varphi} = (\varphi_1, \cdots, \varphi_K)$, the sequence of N target values $\mathbf{y}$ is modelled as a weighted sum of the basis functions plus noise ϵ:

$$\mathbf{y} = f(\mathbf{x}, \mathbf{w}) + \epsilon = \boldsymbol{\varphi}(\mathbf{x})\mathbf{w} + \epsilon \tag{1}$$

where we use the notation $\boldsymbol{\varphi}(\mathbf{x})$ to denote the $N \times K$ matrix with element $\boldsymbol{\varphi}_{i,k} = \varphi_k(x_i)$, and where $\mathbf{w}$ is a vector of K weights.

2.3 Learning basis function weights from data

Given performances in form $(\mathbf{x}, \mathbf{y})$ we use the model in equation (1) to estimate the weights $\mathbf{w}$, which is a straightforward linear regression problem. Under the assumption that the noise ϵ is normally distributed, the maximum likelihood estimation of $\mathbf{w}$ is the least squares solution, that is, the $\mathbf{w}$ that minimises the sum of the squared differences between the predictions $f(\mathbf{x}, \mathbf{w})$ of the model and the target $\mathbf{y}$:

$$\hat{\mathbf{w}} = \operatorname{argmin}_{\mathbf{w}} \|\mathbf{y} - \boldsymbol{\varphi}(\mathbf{x})\mathbf{w}\ \| \tag{2}$$

The simplest approach to find the optimal $\mathbf{w}$ for a data set $D = (\ (\mathbf{x}_1, \mathbf{y}_1), \cdots, (\mathbf{x}_L, \mathbf{y}_L)\ )$, is to concatenate the respective $\mathbf{x}_l$'s and $\mathbf{y}_l$'s in D into a single pair $(\mathbf{x}, \mathbf{y})$, and to find $\hat{\mathbf{w}}_D$ according to equation (2), using $\mathbf{x}$ and $\mathbf{y}$. However, this approach can only be applied when there is a fixed set of basis functions across all pieces, as is the case when only global bases are used.

Another, more general approach is to compute a vector $\hat{\mathbf{w}}_l$ for each performance $1 \leq l \leq L$ according to equation (2). This allows the weight vectors $\hat{\mathbf{w}}_l$ to be of different lengths (as with local bases). In that case, there is no single estimate $\hat{\mathbf{w}}_D$ of the weights for all bases based on the performances in D. The inference of appropriate weights for a new performance can now be regarded as a regression problem given the estimated weight vectors $(\hat{\mathbf{w}}_1, \cdots, \hat{\mathbf{w}}_L)$, and a partition of those weights. In subsection 2.3.1, we describe this approach in more detail.

2.3.1 Prediction with local and global bases

In the case of global bases, once a weight vector $\hat{\mathbf{w}}_D$ has been learnt from a data set D, predictions for a new score $\mathbf{x}$ can be made easily, using equation (1), leaving out the noise term ϵ. First, we construct the matrix of basis functions $\boldsymbol{\varphi}(\mathbf{x})$ from $\mathbf{x}$, and subsequently we apply the dot product of the matrix with the learnt weights $\hat{\mathbf{w}}_D$:

$$\hat{\mathbf{y}}_D = f(\mathbf{x}, \hat{\mathbf{w}}_D) = \boldsymbol{\varphi}(\mathbf{x})\hat{\mathbf{w}}_D \tag{3}$$

In the case of local bases, weight estimation is realised for each piece in D individually, leading to a set of weight vectors $(\hat{\mathbf{w}}_1, \cdots, \hat{\mathbf{w}}_L)$. As described in subsection 2.1.5, local basis functions are 'instantiations' of a basis function 'class'. For example, a score may give rise to three bases of the class *crescendo*, for each of three *crescendo* signs occurring in a score. In this way, we can associate each estimated weight $\hat{w}_j$ with the type of its corresponding basis function c_j. This leads to a set of pairs $(c_j, \hat{w}_j)$, to be used as training data for any regression algorithm, to predict a weight w for a given basis function type c. This approach allows for arbitrarily rich descriptions of basis-functions: rather than characterising a basis function just as a *crescendo*, it might for example be characterised as a *crescendo* following a *piano*, in a minor key context.

3 Experimentation

This section consists of two experiments on different data sets. The first experiment is intended to demonstrate the utility of the model as a method to account for aspects of musical expression. For this experiment, we use a large data set of precisely measured performances by a single professional pianist. The evaluation considers both how well the expressive dynamics variations can be *represented* by the model using various combinations of basis functions, and how well the model, when trained on data generalises to unseen data.

The second experiment demonstrates how the model can be used as an analysis tool, to study differences between the expressive interpretations of different performers. For this we use a smaller data set with loudness values computed from commercial recordings of performances by various famous pianists.

In both experiments, we restrict our attention to expressive dynamics. The main reason for this is pragmatic: dynamics annotations appear more frequently than other types of annotations in the scores we are considering, and therefore models of expressive dynamics serve better to demonstrate the utility of the approach.

3.1 Experiment 1: Representation and prediction of expressive dynamics

The objective of this experiment is to assess how accurately expressive dynamics can be represented and predicted using LBM. In particular, we are interested which (combination) of the basis functions described above is most useful for representation and prediction.

3.1.1 Method

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); 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 (1992) model of dynamics.

We employ two different modelling approaches. In the first, all features are represented by global basis functions (see subsection 2.1.5), and the weights for the bases are estimated all at once by concatenating all performances in the training set, as described in subsection 2.3. In the second scenario, we use global bases to represent the features PIT, GR, and IR, and local bases for DYN. In this case, we use the second weight estimation procedure described in 2.3, in which weights are learnt per piece/performance pair. For the prediction of expressive dynamics for unseen data, support vector regression (Schölkopf and Smola, 2002) was used to estimate weights for unseen data, based on the estimated weights from the training data (cf. subsection 2.3.1).

The total number of model parameters in the global scenario 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 local scenario the number of parameters can either be larger or smaller than in the global scenario, depending on the number of dynamics markings that appear in the piece. In the evaluation, we omit the feature combinations that consist of only GR and IR, since we expect their influence on dynamics to be marginal with respect to the features DYN and PIT.

3.1.2 Data Set

For the evaluation we use the Magaloff corpus (Flossmann et al., 2010) – a data set that comprises live performances of 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 (Moog and Rhea, 1990) 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 note dynamics in the form of MIDI 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.

It is likely that Magaloff used manuscripts as scores, but we are uncertain as to the exact version. To obtain dynamics markings from the scores, we have used the Henle Urtext Edition wherever possible, which explicitly states its intention to stay faithful to Chopin's original manuscripts. The dynamics markings are obtained by optical music recognition from the scanned musical scores (Flossmann et al., 2010).

3.1.3 Goodness-of-fit of the dynamics representation

Table 2 shows a comparison of the observed expressive dynamics with the optimal fit of the model. The goodness-of-fit is expressed in two quantities: r is

	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 3.1 for abbreviations

the Pearson product-moment correlation coefficient, denoting how strongly the observed dynamics and the dynamics values of the fitted model correlate. The quantity R^2 is the coefficient of determination, which expresses the proportion of variance accounted for by the model.Average and standard deviations (in parentheses) of the r and R^2 values over the 154 musical pieces are listed in table 2.

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. We note however, that the r and R^2 values are averaged over pieces, with a considerable standard deviation, and that differences between outcomes may not all be significant.

3.1.4 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

	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 3.1 for abbreviations

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). Note however, that the standard deviation of R^2 is large in most cases.

3.2 Experiment 2: Analysis of expressive dynamics in commercial recordings

The above experiments are all done using the performances of a single performer. In the next experiment, we wish to highlight that the LBM framework can also be used to study differences between performers.

3.2.1 Method

Given the performances of a number of pieces by different performers, we fit an LBM to the expressive dynamics in the performances. The fitted weights are then compared across pieces and performers.

By the nature of the data (see subsection 3.2.2), loudness measurements are only available on a beat level. Moreover the measurements are only approximate,

Piece	Performances
Op. 52	Kissin1998, Pollini1999, Zimerman1987, Horowitz1952/1981, Rubinstein1959, Cherkassky1987, Ashkenazy1964, Perahia1994
Op. 15(1)	Ashkenazy1985, Rubinstein1965, Richter1968, Maisenberg1995, Leonskaja1992, Arrau1978, Harasiewicz1961, Pollini1968, Barenboim1981, Pires1996, Argerich1965, Horowitz1957, Perahia1994
Op. 27(2)	Rubinstein1965, Arrau1978, Kissin1993, Leonskaja1992, Pollini1968, Barenboim1981, Ashkenazy1985, Pires1996, Harasiewicz1961
Op. 28(17)	Sokolov1990, Arrau1973, Harasiewicz1963, Pogorelich1989, Argerich1975, Ashkenazy1985, Rubinstein1946, Pires1992, Kissin1999, Pollini1975

Table 4: Performances used for evaluation

so a fitting a complex model with many basis-functions is likely to capture a lot of measurement noise. For that reason, we use only the DYN basis functions.

We use a local basis fitting approach, in which each piece/performance pair is fitted individually.

3.2.2 Data

Data as precisely measured as that of the Magaloff corpus is not available for other performers – at least not for several performers playing the same pieces of music. We use loudness measurements from commercial CD recordings as an alternative. The data for this experiment has been used earlier by Langner and Goebl (2003). Loudness from the PCM data was calculated using an implementation of Zwicker and Fastl's (2001) psycho-acoustic model, by Pampalk et al. (2002). To reduce the effect of different recording levels, the data was transformed to have zero mean and unit standard-deviation per piece, as in Repp (1999). In addition to the loudness computation, beat-tracking was performed semi-automatically, using *BeatRoot* (Dixon, 2001).

The data set includes multiple performances of four Chopin piano pieces: a Ballade, a Prelude, and two Nocturnes. The performances for each piece are listed in table 4.

3.2.3 Results

In columns 2 and 3 of table 5 (under *meas. vs. fit*), R^2 and r measures are shown per piece, averaged over all performers. Analysis of variance shows that both R^2 and r differ significantly across pieces: $F(3,8) = 30.15$, $p < .001$, and $F(3,8) = 26.51$, $p < .001$, respectively[5]. No effect of performer on goodness-of-fit measures was found.

[5] To meet the assumptions of ANOVA, the data set was restricted to the pianists for which performances of all four pieces are available, namely Pollini, Rubinstein, and Ashkenazy

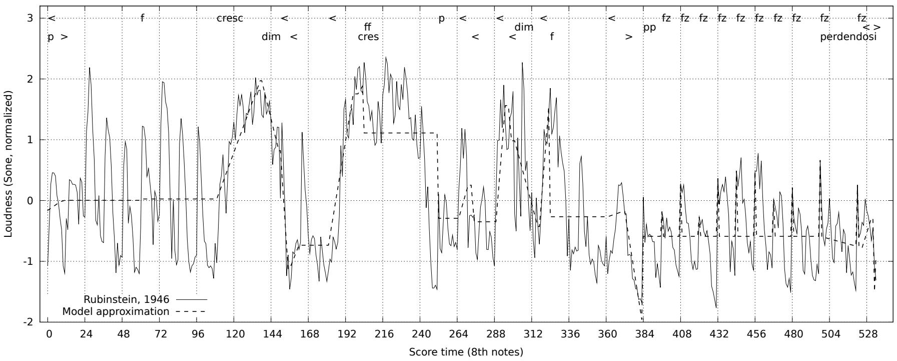


Figure 4: Example of loudness curve and a fitted model; Solid curve: loudness as measured from Rubinstein's performance (1946) of Chopin's Prelude (Op. 28, No. 17); Dashed curve: approximation of the loudness curve by the linear basis model; loudness directives are displayed above the curves

	meas. vs. fit		measurement	residual
Piece	R^2	r	r	r
Op.(No.)	mean (sd)	mean (sd)	mean (sd)	mean (sd)
15 (1)	0.90 (0.04)	0.95 (0.02)	0.88 (0.03)	0.50 (0.11)
27 (2)	0.76 (0.07)	0.87 (0.04)	0.80 (0.03)	0.56 (0.07)
28 (17)	0.66 (0.08)	0.81 (0.05)	0.76 (0.06)	0.61 (0.05)
52	0.86 (0.05)	0.93 (0.03)	0.82 (0.06)	0.48 (0.08)

Table 5: Mean and standard deviation of R^2 and r per piece

Columns 4 (*measurement*) and 5 (*residual*) of table 5 summarise the correlations between the loudness curves of performers. The r values in column 4 are computed on the measured loudness curves. Column 5 contains the r values computed from the residual loudness curves, after the model fit has been subtracted.

The variance of coefficients across pieces appears to be too large to reveal any direct relationships between performers and coefficients, independent of the piece. Within pieces however, significant effects of performer on coefficients are present for the coefficients of some loudness directives. For example, in Op. 52. there is an effect of performer on *ff* coefficients ($F(7, 28) = 3.90,\ p < .005$), and in Op. 28 (No. 17), an effect of performer on *fz* coefficients ($F(9, 90) = 25.75,\ p < .0001$).

4 Discussion

In this section we discuss the results of both experiments described in the previous section. We conclude the section with a brief discussion on how we see the

LBM framework in relation to creative aspects of expressive music performance.

4.1 Discussion of experiment 1

The results presented in experiment 1 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 dynamics variance than the annotations may be surprising, given that annotations are by nature intended to guide dynamics. One may hypothesise that the effect of pitch on dynamics is due to the fact that on a piano different keys must be struck with different intensities to achieve the same sound pressure level (SPL) Goebl and Bresin (2003). However, on the Bösendorfer, the pitches around C5 (midi value 72) produce a higher SPL at the same MIDI velocity than lower pitches. Thus, the pitch effect on dynamics is not a matter of SPL compensation.

Although the data set contains many performances, it is important to realise 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 dynamics may be different for other performers, composers, and musical genres. Specifically, we hypothesise that the fact that pitch effect on dynamics is a consequence of *melody lead*. This phenomenon, which has been the subject of extensive study (see Repp (1996); Goebl (2001)), 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 recognisable by the listener, and may improve the sensation of a coherent musical structure. In many musical genres, the main melody of the music is expressed in the highest voice, which explains the relationship between pitch and dynamics.

This effect is clearly visible in figure 5, which displays observed, fitted, and predicted dynamics for the final measures of Chopin's Prelude in B major (Opus 28, No. 11). In this plot, the velocity 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 velocity of melody notes is substantially higher than the velocity of non-melody notes. This effect is very prominent in the predictions of the model as well. [6]

Although observed and predicted dynamics are visibly correlated, figure 5 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 model parameters have been optimised to performances of a wide range of different pieces, preventing the model from accurately capturing dynamics in individual performances. This suggests a separate treatment of musical pieces with distinct musical characters.

[6] See www.cp.jku.at/research/TRP109-N23/BasisMixer/midis.html for sound examples

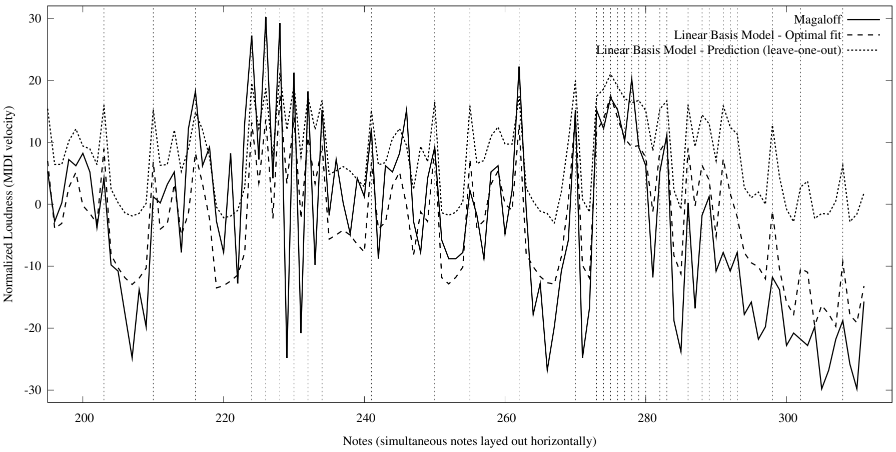


Figure 5: Observed, fitted, and predicted note-by-note dynamics of Chopin's Prelude in B major (Opus 28, No. 11), from measure 16 onwards; Fitting and prediction was done using the global bases DYN+PIT+GR+IR (see section 3.1); Vertical dotted lines indicate melody notes

4.2 Discussion of experiment 2

The results of experiment 2 show that the LBM in combination with DYN basis functions accounts for a large part of dynamics in music performances (66%–90%, depending on the piece; see table 5). The residual loudness after subtracting model fits is substantially less correlated between performers. The remaining correlation is an indication of factors that are not represented by the model. Obvious candidates are pitch, and the number of simultaneously sounding notes.

It is unlikely however, that the described method in its current form will result in clear 'coefficient profiles' of performers, i.e. sets of coefficients that uniquely characterise how a performer interprets annotations. Many decisions on how to interpret annotations will depend on the context of the annotation and on musical understanding of a level that is not easy to capture in a simple mathematical model.

Nevertheless, LBM can be a useful tool to compare interpretations of different performers for a particular piece or musical fragment. It provides estimates of how (strongly) each annotation has shaped the dynamics of the performance. Although the model provides only an approximation of the performed dynamics, these estimates can often be meaningfully compared across performers.

4.3 LBM and creativity in musical expression

The LBM framework presented in this paper is an attempt to account for musical expression in a rather simple manner, namely as a weighted sum of scoredetermined basis functions. One may wonder whether such a model leaves any room for other aspects of musical expression, such as the performer's expressive intentions, especially the affective information she wishes to transmit to the listener.

Theoretically speaking, since the models factorise musical expression into weights and basis functions, any expressive information should be captured either in one, or in the other. In this paper, we have treated the basis functions as a fixed part of the model, and the weights as parameters to be fit to data. One (admittedly simplistic) way to conceive of the artistic freedom of performers, is to regard the weights as a "palette" with which performers "colour" the performance differently, depending on which basis functions receive high weights. This idea has been proposed for a system of rules for musical expression (Bresin and Friberg, 2000).

It is also plausible that expressive variation due to affective intentions, or individual performer style, may be better modelled by adapting the basis functions. Possibly, the shape of a *crescendo*, or a *ritardando* may be affect-specific, or even performer specific. Some evidence for performer-specific final ritard shapes has been found (Grachten and Widmer, 2009). To learn the shape of basis functions from data however, constraints must be imposed to avoid that the model is under-determined.

It should be stated clearly however, that the LBM framework is intended as a tool for analysing artifacts, rather than the process that led to these artifacts. Analogously, the process of creating a new performance by using LBM, should not be seen as modelling a cognitive process, let alone a creative process. Whether a performance created by LBM could be regarded as creative by a human listener is a philosophical question. We adhere to the view stated by Widmer et al. (2009), that creativity is in the eye of the beholder. It is even conceivable that the creativity is not just a (subjective) characteristic of the performance, but also of the listener's interpretation, by which she construes a novel and unconventional performance as an enjoyable one.

5 Conclusions and future work

The work presented in this paper corroborates the growing insight in music performance research, that even if musical expression is a highly complex phenomenon, it is by no means fully unsystematic. We have described a linear basis modelling framework to account for expressive variations in music performance. Several types of basis functions were discussed. Using a relatively small set of basis functions, it is possible to account for over 45% variance in the dynamics of Magaloff's performances of Chopin piano works. Prediction of performances using a model trained on performance data unsurprisingly yields lower values, but substantial positive correlations are still observed.

As an analytical tool, we have used the framework to quantise performance differences between performers, and pieces. Results indicate that the variance across pieces is too large to identify performer-specific expressive style, but within pieces, some performer-specific expressive effects were identified.

The LBM framework can be extended in two important ways. Firstly, we believe the framework is well-suited to a probabilistic approach, in which prior information on the distribution of weights is combined with estimates obtained from performance data. Secondly, a strong limitation of the current model is that basis functions must be defined manually. Dictionary learning techniques developed in the field of sparse coding may be used to learn basis-functions from performances.

Acknowledgements

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 the data of her late husband's performances for our research. For this research, we have made extensive use of free software.

References

Bresin, R. and Friberg, A. (2000). Emotional coloring of computer-controlled music performances. *Computer Music Journal*, 24(4):44–63.

Canazza, S., De Poli, G., Drioli, C., Rodá, A., and Vidolin, A. (2004). Modeling and control of expressiveness in music performance. *Proceedings of the IEEE*, 92(4):686–701.

Canazza, S., De Poli, G., and Rodá, A. (2002). Analysis of expressive intentions in piano performance. *Journal of ITC Sangeet Research Academy*, 16:23–62.

Clarke, E. F. (1988). Generative principles in music. In Sloboda, J., editor, *Generative Processes in Music: The Psychology of Performance, Improvisation, and Composition*. Oxford University Press.

Dixon, S. (2001). Automatic extraction of tempo and beat from expressive performances. *Journal of New Music Research*, 30(1):39–58.

Flossmann, S., Goebl, W., Grachten, M., Niedermayer, B., and Widmer, G. (2010). The Magaloff Project: An Interim Report. *Journal of New Music Research*, 39(4):369–377.

Goebl, W. (2001). Melody lead in piano performance: expressive device or artifact? *Journal of the Acoustical Society of America*, 110(1):563–572.

Goebl, W. and Bresin, R. (2003). Measurement and reproduction accuracy of computer controlled grand pianos. *Journal of the Acoustical Society of America*, 114(4):2273–2283.

Good, M. (2001). MusicXML for notation and analysis. In Hewlett, W. B. and Selfridge-Field, E., editors, *The Virtual Score: Representation, Retrieval,* *Restoration*, volume 12 of *Computing in Musicology*, pages 113–124. MIT Press, Cambridge, MA.

Grachten, M. (2006). *Expressivity-aware Tempo Transformations of Music Performances Using Case Based Reasoning*. PhD thesis, Pompeu Fabra University, Barcelona, Spain. ISBN: 635-07-094-0.

Grachten, M. and Widmer, G. (2009). Who is who in the end? recognizing pianists by their final ritardandi. In *Proceedings of the 6th Sound and Music Computing Conference (SMC)*, Porto, Portugal.

Grachten, M. and Widmer, G. (2011). Explaining expressive dynamics as a mixture of basis functions. In *Proceedings of the Eighth Sound and Music Computing Conference (SMC)*, Padua, Italy.

Krebs, F. and Grachten, M. (2012). Combining score and filter based models to predict tempo fluctuations in expressive music performances. In *Proceedings of the Ninth Sound and Music Computing Conference (SMC)*, Copenhagen, Denmark.

Langner, J. and Goebl, W. (2003). Visualizing expressive performance in tempo-loudness space. *Computer Music Journal*, 27(4):69–83.

Moog, R. A. and Rhea, T. L. (1990). Evolution of the Keyboard Interface: The Bösendorfer 290 SE Recording Piano and the Moog Multiply-Touch-Sensitive Keyboards. *Computer Music Journal*, 14(2):52–60.

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

Palmer, C. (1996). Anatomy of a performance: Sources of musical expression. *Music Perception*, 13(3):433–453.

Palmer, C. (1997). Music performance. *Annual Review of Psychology*, 48:115–138.

Pampalk, E., Rauber, A., and Merkl, D. (2002). Content-based organization and visualization of music archives. In *Proceedings of the 10th ACM International Conference on Multimedia*, pages 570–579. ACM.

Parncutt, R. (2003). *Perspektiven und Methoden einer Systemischen Musikwissenschaft*, chapter Accents and expression in piano performance, pages 163–185. Peter Lang, Germany.

Repp, B. H. (1996). Patterns of note onset asynchronies in expressive piano performance. *Journal of the Acoustical Society of America*, 100(6):3917–3932.

Repp, B. H. (1999). A microcosm of musical expression: II. Quantitative analysis of pianists dynamics in the initial measures of Chopin's Etude in E major. *Journal of the Acoustical Society of America*, 105(3):1972–1988.

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

Schölkopf, B. and Smola, A. J. (2002). *Learning with Kernels*. MIT Press.

Timmers, R.and Ashley, R., Desain, P., Honing, H., and Windsor, L. (2002). Timing of ornaments in the theme of Beethoven's Paisiello Variations: Empirical data and a model. *Music Perception*, 20(1):3–33.

Tobudic, A. and Widmer, G. (2003). Playing Mozart phrase by phrase. In *Proceedings of the Fifth International Conference on Case-Based Reasoning (ICCBR-03)*, number 2689 in Lecture Notes in Artificial Intelligence, pages 552–566. Springer-Verlag.

Todd, N. (1992). The dynamics of dynamics: A model of musical expression. *Journal of the Acoustical Society of America*, 91:3540–3550.

Widmer, G., Flossmann, S., and Grachten, M. (2009). YQX plays Chopin. *AI Magazine (Special Issue on Computational Creativity)*, 30(3):35–48.

Zwicker, E. and Fastl, H. (2001). *Psychoacoustics: Facts and Models*. Springer-Verlag, Berlin, 2nd edition.