International Symposium on Performance Science
ISBN 978-94-90306-02-1

A method to determine the contribution of annotated performance directives in music performances

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

[1] Department of Computational Perception, Johannes Kepler University, Linz, Austria
[2] Austrian Research Institute for Artificial Intelligence, Vienna, Austria

Interpreting notated music and performing it expressively is a complex skill that requires years of practice. In the quest for understanding this phenomenon, a question that arises naturally is to what degree performance directives annotated in the score affect expressive variations of tempo and loudness. Computational models of musical expression typically focus on musical structure and do not explicitly take into account annotated performance directives. The objective of the method presented here is to determine the degree to which loudness directives can account for expressive variations in loudness as measured from performances. To this end, we represent loudness directives by mathematical functions and use these to approximate measured loudness curves. This approximation yields coefficient values that represent how strongly each directive is reflected in the performance. Furthermore, the residual loudness curve after subtracting the model fit provides a clearer view on other, non-explicit factors that influence expressive loudness variations.

Keywords: empirical musicology; expression modeling; music performance; performance directives; computational analysis

Interpreting notated music and performing it expressively is a complex skill that requires years of practice. Many empirical studies exist that aim to elucidate the process of expressive performance. A factor that is considered to be of major influence on expressive variations in performance tempo and loudness is the structural interpretation of the music (Clarke 1988). Several computational models have been proposed to hypothesize how structural aspects of music explain expressive performance (Todd 1992, Parncutt 2003).

Composers often give hints as to appropriate loudness and tempo by annotating scores with performance directives, such as *crescendo* and *accelerando*. Such directives are typically not a formal part of computational models of expressive performance. A possible explanation for this is that performance directives are often regarded as incidental to pitch and temporal structure, the latter two being considered the essence of notated music. In some cases this may be true, for example where the expressive hints of composers have led to a performance practice that marks a particular musical genre. In other cases, composers may place highly specific, non-obvious markings to ensure the performance achieves the intended effect. Rosenblum (1988) offers a discussion of the interpretation of expressive markings in composers' works.

In this paper, we present a novel method to disentangle the interpretation of explicitly written loudness directives from non-explicit forms of loudness variation. On the one hand, this allows us to determine how such directives are interpreted by performers; on the other, it may provide a clearer view of expressive interpretation beyond the written directives. The method follows the common intuition that musical expression consists of a number of individual factors that jointly determine what the performance of a musical piece sounds like (Palmer 1996). The goal is then to identify which factors can account for expressive dynamics and to disentangle their contributions to the loudness of the performance.

METHOD

The objective of the presented method is to determine the degree to which loudness directives can account for expressive variations in loudness measured from performances. To this end, we represent each loudness directive by mathematical functions, henceforth called *basis functions*. Each basis function represents loudness as a function of time, over the scope of its corresponding directive. The functions are weighted and summed to approximate loudness curves of performances, as illustrated in Figure 1.

We distinguish between three categories of loudness directives. The first category, *constant*, represents markings that indicate a particular loudness character for the length of a passage. This category includes the familiar markings (*p*, *f,* etc.) and adjectives such as *dolce* and *leggiero*. The second category, *impulsive*, indicate a sudden and brief change of loudness, such as *fp* and *sf (sforzando)*. The third category, *gradual*, contains directives that indicate a gradual change from one loudness level to the other, such as *(de)crescendo*, but also metaphorical descriptors of dynamic evolution, such as *perdendosi* and *smorzando*.

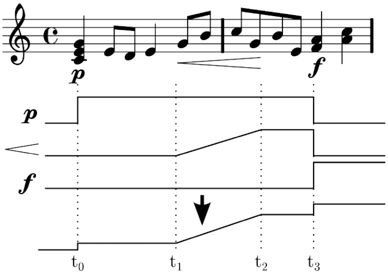

Figure 1. Example of basis functions representing performance directives.

We assign a particular basis function to each category. The markings of the constant category are modeled as step functions that have value 1 over the affected passages, and 0 elsewhere. Impulsive directives are modeled by unit impulse functions, which have value 1 at the time of the directive and 0 elsewhere. Lastly, gradual directives are modeled as a combination of a ramp and a step function, which is 0 until the start of the directive, linearly changes from 0 to 1 between the start and the end of the indicated range of the directive (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 directive.

With this mapping of performance directives to functions, sequences of directives, as read from a musical score, can be translated to a set of basis functions. Finding the weighting coefficients that make the weighted sum of these basis functions approximate a measured loudness curve as closely as possible is an example of linear regression, a well-known algebraic problem. The optimal coefficients can be found using least-squares minimization. This yields one coefficient for each loudness directive, representing the strength of that directive in the performance.

Data

We test the above method on a set of performances of Chopin's piano music, performed by a number of famous pianists. The data have been extracted from CD recordings and used earlier by Langner and Goebl (2003). Perceived loudness calculation from the audio data was done using Zwicker and Fastl's (2001) psychoacoustic model. To eliminate any effects of recording quality, the data was transformed to have zero mean and unit standard deviation per piece, as in Repp (1999). The data set includes multiple performances of four Chopin piano pieces, as listed in Table 1.

Table 1. Performances used for evaluation of the method.

Piece	*Performances*
Op. 15 (1)	Ashkenazy 1985, Rubinstein '65, Richter '68, Maisenberg '95, Leonskaja '92, Arrau '78, Harasiewicz '61, Pollini '68, Barenboim '81, Pires '96, Argerich '65, Horowitz '57, Perahia '94
Op. 27 (2)	Rubinstein '65, Arrau '78, Kissin '93, Leonskaja '92, Pollini '68, Barenboim '81, Ashkenazy '85, Pires '96, Harasiewicz '61
Op. 28 (17)	Sokolov '90, Arrau '73, Harasiewicz '63, Pogorelich '89, Argerich '75, Ashkenazy '85, Rubinstein '46, Pires '92, Kissin '99, Pollini '75
Op. 52	Kissin '98, Pollini '99, Zimerman '87, Horowitz '52/'81, Rubinstein '59, Cherkassky '87, Ashkenazy '64, Perahia '94

RESULTS

The linear basis model was fitted to the measured loudness curves of the performances. An example is shown in Figure 2. The goodness-of-fit is quantified in two measures: the coefficient of determination (R^2), expressing the proportion of variance in the measured loudness curves explained by the fitted model, and the correlation coefficient (r), expressing linear dependence between model fit and measurement. In columns 2 and 3 of Table 2 (*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: $F_{3,8}=30.15$, $p<0.001$, and $F_{3,8}=26.51$, $p<0.001$, respectively. [*Note.* To avoid an unbalanced setup, only the performance of Pollini, Rubinstein, and Ashkenazy were used for the ANOVA.] No effect of performer on goodness-of-fit measures was found. Columns 4 (*Measurement*) and 5 (*Residual*) of Table 2 summarize 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 simple 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 a performer effect on *ff* coefficients ($F_{7,28}=3.90$, $p<0.005$) and in Op. 28 (No. 17) on *fz (sforzando)* coefficients ($F_{9,90}=25.75$, $p<0.0001$).

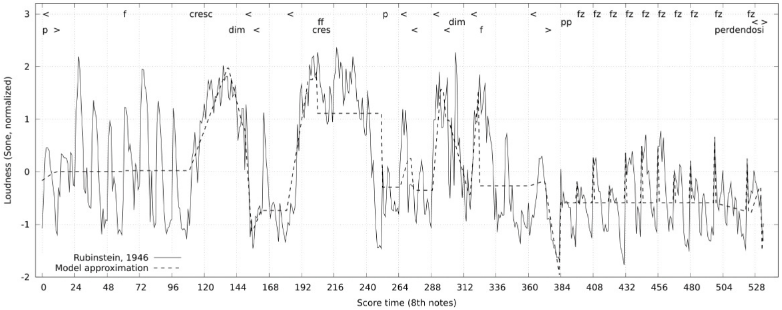

Figure 2. Example of a loudness curve and a fitted model. Solid line=loudness as measured from Rubinstein's performance (1946) of Chopin's Prelude, Op. 28, No. 17; Dashed line=approximation of the loudness curve by the linear basis model; loudness directives are displayed above the curves, where < and > denote "hairpin" crescendi/diminuendi, and *cresc.* and *dim.* denote longer range crescendi/diminuendi.

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

	Meas. vs. Fit		*Measurements*	*Residual*
Piece	R^2	*r*	*r*	*r*
Opus (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)

DISCUSSION

The results show that the proposed model accounts for a large part of loudness variations in music performances. 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 sounding notes at a specific time. As shown elsewhere (Grachten and Widmer 2011), the model also accommodates for such factors in the form of basis functions.

It is unlikely that the described method in its current form will result in clear "coefficient profiles" of performers—i.e. sets of coefficients that uniquely characterize how a particular performer interprets loudness directives. Many decisions on how to interpret directives will depend on context and on musical understanding of a level that is not easy to capture in a simple mathematical model. Nevertheless, the linear basis model can be a useful tool to compare interpretations of different performers for particular pieces or excerpts. It provides estimates of how strongly each loudness directive has shaped the loudness of a performance. Although one should keep in mind that the model gives only an approximation of performed loudness, these estimates can often be compared across performers in a meaningful way.

Acknowledgments

This research is supported by the Austrian Research Fund (FWF Z159).

Address for correspondence

Maarten Grachten, Department of Computational Perception, Johannes Kepler University, Altenbergerstr. 69, Linz 4040, Austria

References

Clarke E. F. (1988). Generative principles in music. In J. Sloboda (ed.), *Generative Processes in Music* (pp. 1-26). Oxford: Oxford University Press.

Grachten M. and Widmer G. (2011). *Explaining Musical Expression as a Mixture of Basis Functions*. Manuscript under review.

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

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

Parncutt R. (2003). Accents and expression in piano performance. In F. Niemöller and B. Gätjen (eds.), *Perspektiven und Methoden einer Systemischen Musikwissenschaft* (pp. 163-185). Köln, Germany: Peter Lang.

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,* pp. 1972-1988.

Rosenblum S. P. (1988). *Performance Practices in Classic Piano Music*. Bloomington, Indiana, USA: Indiana University Press.

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

Zwicker E. and Fastl H. (2001). *Psychoacoustics* (2nd ed.). Berlin: Springer-Verlag.