Modelling Musical Dynamics

Axel Berndt, Tilo Hähnel
Department of Simulation and Graphics
Otto-von-Guericke University
Universitätsplatz 2
39106 Magdeburg, Germany
{aberndt|tilo}@isg.cs.uni-magdeburg.de

ABSTRACT

This paper deals with dynamics, i.e. loudness, in music. We developed models to describe and recreate dynamics properties of human musicians' expressive performances within a performance system. Their diversity and variability bears a particular challenge to the flexibility of such models. Therefore, our approach is based on a multi-layered discrimination of different dynamics aspects. Each of which includes certain degrees of freedom to flexibly define different shapings. For evaluation we compare these models with human performed dynamics.

Categories and Subject Descriptors

I.6.5 [**Simulation and Modeling**]: Model Development; J.5 [**Arts and Humanities**]: *music, performing arts*

General Terms

Music, Expressive Performance

Keywords

Dynamics

1. INTRODUCTION

Musical dynamics touches everything concerning loudness in music. When musicians use dynamics to intensify the expressivity of their performance, they have to decide when and why they play loud, soft, perform a crescendo or an accent, and so on. These questions are, in fact, crucial and entail complex loudness shapings. Generally, expressive dynamics is more than a sequence of single events. In nearly every performance an interplay of several different dynamics layers creates permanent loudness changes.

How to model this within a performance system? Some rule-based approaches derive such properties from aspects of the compositional structure like the degree of dissonance, the pitch of a tone [4], or the phrasing [12]. Others model dynamics in accordance with certain timing aspects [13, 16], construct it within a comprehensive mathematical music theory [6], or apply machine-learning techniques to capture human musicians' individual performance styles [15]. Most of these approaches are strongly oriented towards performance principles that were established in the Romantic era. These do not hold generally to approach expressive dynamics in different performance styles, though [10]. Such a generalization claim leads to a more fundamental question: How are dynamics, and especially dynamics transitions, shaped and what degrees of freedom exist?

This paper traces our research into this question. It is embedded in a series of detailed investigations that covers the 'three pillars' of expressive music performance, namely dynamics, timing [2], and articulation [5]. Theorization and detailed analyses of human performances lead to deep insights and new approaches to mathematically model performance properties and implement them within a performance rendering engine. This engine is developed as a tool for musicological performance research (e.g., for the reconstruction of historically informed performance practices) and for music production in general. It further implements several techniques that allow the interactive combination of different performance styles which is particularly interesting in the games scoring context [1]. The developments and findings described in this work are highly beneficial for the quality of the interaction-driven performance transitions. A further application scenario is the performance of computer-generated music.

A major concern when analyzing human performances is to identify variability within dynamics processes. A crescendo is not always performed in the same way, the loudness of forte differs from instrument to instrument, and so on. For this purpose, Baroque music proved to be particularly beneficial. In this stylistic era all types of dynamics were established and performative decisions (which affect the shaping of dynamics processes) feature a large amount of individual freedom. Nearly every dynamics-related decision is made by the musician and incorporates prior knowledge of common practices. Over succeeding stylistic eras dynamics became more and more a subject of compositional decisions and was notated more detailed in the score. Moreover, the rigid rule of dynamics that have to mimic phrase arches was not yet common in the Baroque.

The remainder of this paper is structured as follows. The theoretical basis to formalize dynamics is introduced in Sec-

AM'10, September 15–17, 2010, Piteå, Sweden.

tion 2. Section 3 develops the models to capture dynamics features and their variable shaping. These are complemented in Section 4 showing that continuous dynamic transitions played by human musicians are neither linear nor generalizable.

2. DEFINITIONS

In its broadest sense musical dynamics comprizes all aspects of loudness in music. This is not one homogeneous class of features but a conglomerate of several feature classes. These have different characteristics and are widely independent of each other. Their only common property is that they affect the loudness of musical events. Their most striking difference is the temporal extent of their effect. In this respect, we distinguish *macro dynamics* and *micro dynamics*, i.e., temporally large-scale and small-scale dynamics.

This Section describes their formalization. It is basis for both, the analysis of human musician performances and the implementation within a music engine to render expressive performances in the MIDI format.

2.1 Macro Dynamics

Macro dynamics is what is traditionally known as dynamics in music and indicated by instructions like *piano*, *mezzoforte*, *forte*, *crescendo*, and *decrescendo*. Two types of instructions can be distinguished: (i) those that discretely set a loudness level and (ii) those that dictate continuous dynamics transitions. The temporal range of an instruction is terminated by its successor and comprizes multiple measures, phrases, even up to the whole piece of music. It is rarely less than one measure.

In a polyphonic musical setup dynamics do not have to be equal in all parts. For instance, those parts that perform a melodic task may be set louder than their accompaniment. This may change later on in the piece. Consequently, macro dynamics is formally a sequence of dynamics instructions I_j, dedicated to one or more musical parts. We call this sequence a *dynamics map* M_D in correspondence to the well-known tempo map paradigm.

$$M_D = (I_0, I_1, ..., I_i)$$

One such instruction I_j has the form

$$I_j = (d_j, v_{1_j}, v_{2_j}, \text{shape}_j), j \in [0, i]$$

with d_j the tempo-independent musical date of the instruction (score position or MIDI ticks), the two loudness values v_{1_j} (initial loudness) and v_{2_j} (target loudness), and the so-called *shape term* (shape_j). Thereby, instruction I_j describes a continuous loudness transition from v_{1_j} to v_{2_j} in the timeframe $[d_j, d_{j+1})$ with shape_j. The construction of this shape is detailed in Section 3.1. If I_j is the last instruction in M_D its timeframe terminates with the end of the musical piece. When working with MIDI sequences this is the last event in the sequence, usually the EndTrack event.

For the technical implementation the loudness values v_{1_j} and v_{2_j} must be declared numerically. The KeyOn velocity in the MIDI standard, for instance, allows only integers from 0 (muted) to 127 (loudest). A different convention is commonly used in music notation. Here loudness is indicated by strings like *ff*, *f*, *mf*, *p*, *pp*, etc. Our implementation supports both conventions so as to make the manual editing over a high-level description language (XML-based) more intuitive. Therefore, we introduce a lookup table as header to the dynamics map which defines the mapping from string to numerical value. An example is shown in Table 1.

String	*ppp*	*pp*	*p*	*mp*	*mf*	*f*	*ff*	*fff*
MIDI vel.	2	36	48	64	83	97	111	125

Table 1: An example lookup table for the mapping of dynamics instructions to MIDI velocity values.

Different instruments can further differ with regard to their loudness scope [8]. Recorders, for instance, are relatively quiet and have very little scope to vary. Brass instruments, by contrast, feature an ample scope upwards (louder) but playing quiet is relatively hard. As a consequence, a recorder *forte* would differ significantly from a brass *forte*. As far as the sampler or synthesizer in use does no such scaling it has to be done by the software. In fact, it is relatively easy to formally encapsulate this variance. As each instrument/part has its own dynamics map it also has its own lookup table which can define different velocity values. The lookup table can also be used to compensate the differing velocity interpretation of samplers and synthesizers [3].

2.2 Micro Dynamics

Micro dynamics add fine variations to the underlying macro dynamics. The temporal extent of micro dynamics features does not exceed the length of one measure, in contrast to macro dynamics. Two classes of micro dynamics features have to be distinguished: *metrical emphasis* and *articulation*. Articulation is a very comprehensive aspect of music performance that affects all facets of the forming of each single notes, namely envelope, duration, timbre, intonation, and loudness. Because of its complexity, it is traditionally treated as a separate concept, apart from dynamics but related to it. So do we; the further text will omit articulation. We address this aspect exclusively and more detailed in [5].

Time signature and musical style often suggest specific metrical emphasis or accentuation schemes that recur measure-wise. It reflects the distinction between weak and strong beats, the perceptual grouping of pulses according to a metrical structure that is defined by the time signature. As already described for macro dynamics, the metrical emphasis can as well differ from part to part and can change during the course of the piece. Therefore, all metrical emphasis instructions E_k of a part are organized in an *emphasis map* M_E.

$$M_E = (E_0, E_1, ..., E_n)$$

with

$$E_k = (d_k, \text{scope}_k, (S_k^0, S_k^1, ..., S_k^l)), k \in [0, n]$$

Such an emphasis instruction E_k defines the date d_k from when on it is applied to all following measures up to d_{k+1} or the end of the piece. It furthermore defines a sequence of emphases $S_k^{0...l}$, this is the actual emphasis scheme (detailed in Section 3.2), and a dynamic scope (scope_k) to scale

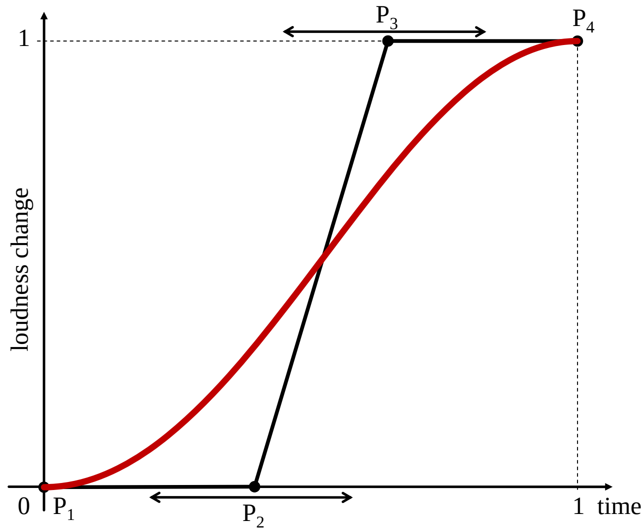

Figure 1: (De-)Crescendi are modelled by cubic Bézier curves. Coordinates x_2 and x_3 are variable.

the accentuation intensity of the scheme. For the MIDI implementation this has to be an integer value in the interval $[0, 127]$.

3. SHAPING DYNAMICS

The previous Section laid the formal basis to describe dynamics features. This Section will now look closely at the modelling aspects.

3.1 (De-)Crescendo

A (macro) dynamics instruction $I_j = (d_j, v_{1_j}, v_{2_j}, \text{shape}_j)$ defines a continuous loudness transition from v_{1_j} to v_{2_j} in the time frame $[d_j, d_{j+1})$. Terraced dynamics can be considered as a special case where v_{1_j} is equal to v_{2_j}. Otherwise, a substantial *crescendo* ($v_{1_j} < v_{2_j}$) or *decrescendo* ($v_{1_j} > v_{2_j}$) is indicated.

In human musicians' performances these will rarely feature a linear behaviour for several reasons. The linear transition appears mechanical and aimless whereas the loudness transition of a human musician seems more determined. Usually, the listener can anticipate the target (the end) of a human performed *(de-)crescendo*. A linear transition would furthermore feature indifferentiabilities, i.e. kinks, at the connections to the previous and following dynamics instruction. These are not necessarily problematic as far as the dynamics change is rendered note-wise (each note with its own loudness) and as these notes are clearly separated from each other (e.g., by staccato or strongly accentuated articulation). However, within a continuous sound stream (held tones, legato play) these kinks are clearly audible just like choppy animations in the visual domain. For more organic animations aspects of inertia have to be considered. Therefore, the animation speed is usually regulated by sigmoidal functions that ensure g1 continuous connections at the beginning and end of animation sections.

As dynamics changes in human-performed music are subject to similar physical factors we also apply sigmoidal characteristics. These are modelled by cubic Bézier curves with Σ-shaped control polygons (see Figure 1). It is defined in the unit square and then scaled to the actual temporal and dynamic extent that is given by the other parameters of the dynamics instruction. Thus, the shape of the transition is defined independent of its scaling. The control points are P_1, P_2, P_3, and P_4 of which P_1 and P_4 are bound to coordinates $(0, 0)$ and $(1, 1)$. The ordinates $y_2 = 0$ and $y_3 = 1$ are fixed, too, whereas x_2 and x_3 are variable in $[0, 1]$. This simplifies the polynomial description of the curve.

$$x(t) = (3x_2 - 3x_3 + 1)t^3 + (-6x_2 + 3x_3)t^2 + 3x_2 t$$

$$y(t) = -2t^3 + 3t^2$$

In our implementation the user does not set x_2 and x_3 directly. Instead, we introduce two descriptors which may be more intuitive: *straightness* (s) and *protraction* (p). These are then converted into x_2 and x_3 values. Thus, the shape of a dynamics transition is defined as

$$\text{shape}_j = (s_j, p_j)$$

The straightness parameter ($s \in [0, 1]$, float) sets the markedness of the S-curvature. The linear shape is created by $s = 0$. The strongest possible curvature is achieved by $s = 1$.

The protraction ($p \in [-1, 1]$, float) introduces a further distortion. It describes a tendency whereby the majority of the loudness transition is made relatively soon ($-1 \leq p < 0$), relatively late ($0 < p \leq 1$), or evenly balanced ($p = 0$).

For $p = 0$ the x-coordinates of P_2 and P_3 are set as follows.

$$x_2 = s \text{ and } x_3 = 1 - s$$

For $p \neq 0$ the straightness-related shifting of x_2 and x_3 has to be scaled to the remaining interval which is no longer equal for both. The conversions are done by the following formulas.

$$x_2 = s + \left(\frac{|p| + p}{2 \cdot p} - \frac{|p| \cdot s}{p} \right) \cdot p$$

$$x_3 = 1 - s + \left(\frac{p - |p|}{2 \cdot p} + \frac{|p| \cdot s}{p} \right) \cdot p$$

The transition can be rendered into the onset/attack loudnesses of the notes. This rough discretization corresponds to a note-wise terraced dynamics. It is sufficient for struck instruments (piano, harp etc.) and usually also suffices for temporally close, i.e. fast, note sequences. However, most wind and string instruments, as well as the human voice, are able to change their loudness even during the sounding tone. We call this phenomenon *sub-note dynamics*. In the MIDI format this can be done through the channelVolume controller. Some samplers and synthesizers offer specialized controllers therefore. The software sampler Vienna Instruments [14], for instance, has the so-called Velocity-Cross-Fade controller.

A continuous loudness change still has to be discretized into a sequence of controller messages. The resolution of this discretization is set by an additional parameter in the shape term, subNoteDynRes.

$$\text{shape}_j = (s_j, p_j, \text{subNoteDynRes}_j)$$

This third parameter defines the step width for t when tracing the Bézier curve by MIDI controller messages. It is defined in the interval $(0, 1]$. E.g., subNoteDynRes = 0.1 triggers 10 controller messages along the Bézier curve, 0.01 triggers 100; the smaller the more. A higher density of messages produces smoother results. However, if the controller value does not change (usually for numerical reasons, MIDI controller values are integer), the message is redundant and causes unnecessary MIDI traffic. This is easily avoided in the implementation by filtering consecutive controllers with equal values or dates.

3.2 Metrical Emphasis Scheme

Amongst the date and scope parameters an emphasis instruction E defines a sequence of emphases $(S^0, S^1, ..., S^l)$. This sequence constitutes the *metrical emphasis scheme* that is repeatedly applied to all measures the instruction covers. Hence, an emphasis scheme describes the loudness contour of one measure which is then added upon the underlying macro dynamics. More precisely, an emphasis scheme defines the contour of the deviations from the underlying macro dynamics. These can also be negative.

An emphasis S^m represents one segment of the scheme.

$$S^m = (b_m, e_{1_m}, e_{2_m}, e_{3_m}), m \in [0, l]$$

The segment begins with beat b_m (for the first beat in the bar $b_m = 1$, second beat $b_m = 2$, etc.) and ends with b_{m+1} or the end of the bar. Hence, in a 3/4-time signature any emphasis with $b_m \geq 4$ will be ignored since the next bar begins already at the fourth beat. Generally, b_m is a floating-point value so that an emphasis can be defined at any position within the bar (e.g., $b_m = 2.5$ designates beat 'two-and').

The emphasis, or accentuation, at bar position b_m is set by e_{1_m} (floating-point value in $[-1, 1]$). $e_{1_m} = 1$ sets the strongest accentuation, 0 causes no deviation from the basic loudness, -1 indicates maximal reserve. These values are scaled by the dynamic scope (defined in E) to the actual loudness deviations which then only have to be added to the macro dynamics. However, if basic loudness plus/minus dynamic scope exceeds or undercuts the MIDI velocity range, it is automatically scaled down. This ensures a differentiated accentuation as far as possible. It furthermore matches a situation in reality: Accentuation is less pronounced at the borders of the dynamic ambit.

Up to now, only discrete emphasis points have been set (b_0 to b_l). To cover the whole segment $[b_m, b_{m+1})$ a transitioning course has to be defined that is applied to the interval (b_m, b_{m+1}). Therefore, we introduce the emphasis values e_{2_m} and e_{3_m}. Both are optional. Without them a constant emphasis of e_{1_m} is set on the whole segment. If the notes between the major beats shall be set at a different emphasis level, e_{2_m} is used. This still produces a constant emphasis but it may suffice for most situations in musical practice.

In contrast to the long macro crescendi and decrescendi that usually involve a fair quantity of notes a segment of an emphasis scheme covers only very few notes and only a very small dynamic range. Nonetheless, their accentuations do not have to be equal/constant. They may, for instance, transition to the emphasis of the following beat or lead pickup-like to the next bar. This is a monotonous behaviour. Therefore, linear functions do well for approximation. In fact, any curved characteristics, such as described in the previous Section for the crescendo/decrescendo, are exaggerated here. Differences between linear and sigmoidal shapes are scarcely perceptible. Thus, we apply linear transitions. These are defined by adding parameter e_{3_m}. The emphasis will now run from e_{2_m} to e_{3_m} in the interval (b_m, b_{m+1}).

4. MEASURING DYNAMICS

Dynamics can be manifold, as the preceding Sections already stated. Regarding the analysis of dynamics Baroque music has several advantages. Continuous dynamics transitions are to a high degree an individual decision of the performer. Unlike in later styles dynamics are rarely annotated but one of the most important expressive tools. The rules that are described in treatises [11, 9] are rather guidelines than strict, describing possible ways to decide whether or not a transition can be applied.

On the other hand, treatises include distinct rules for beat emphases that parallel the hierarchy of beats in a measure. To get empirical data for the intensity of metrical accents one of the most prominent dance-movements, the minuet, was analyzed. The advantage of dance movements like the minuet is that they are repeated several times, thereby increasing the quality of the data collected.

Sub-note dynamics can be observed within long tones that are not accompanied and disrupted by other instruments. Here a solo-piece is more useful.

Generally, the situation is further complicated by the fact that loudness is a subjective impression, which depends on individual physiological thresholds; and the loudness of instruments differs also with respect to pitch and tone color [7]. Hence, research concerning absolute decibel values can only support rules of thumb. Nevertheless it is possible to approximate a curve for normalized data and analyze proportional differences of beat accents.

Consequently, the following conditions were made: (i) a stylistic homogeneity ensured by the selection of similar compositions, (ii) the focus on experts in historically informed performance, for they know the treatises, (iii) use of pieces that include repeated sections, and (iv) the analyses of specific movements and sections, in particular regarding dynamical aspects.

4.1 Methodology

The analyses included live- and studio-recordings. Apart from the experimental recordings all pieces were composed by Georg Philipp Telemann. Except for one studio recording the same orchestra accompanied different soloists. All recordings had the same resolution of 44.1kHz and 16bit. Each tone was represented by its maximum decibel value. To avoid mismatches of loudness values in long crescendo notes the last third of long tones was not taken into account. All loudness values were normalized; in the micro dynamic task the decibel values were first computed as deviations from the mean loudness per measure and normalized afterwards. The compositions used in the analyses are as follows:

Recorder concert: Large crescendi were performed in the second movement (bars 13–16 and 191–194) of the *Concert in C Major for Recorder, String Orchestra and Cembalo*. It was recorded live twice during the final round of the 5th international Telemann competition for historical woodwinds in Magdeburg (TC). The accompanying ensemble played two times with different soloists. The analysis further includes one studio recording of the same orchestra accompanying another soloist.

Flute concert: In the same round of the TC the *Concert in D Major for Flute and String Orchestra* was recorded to analyze a short crescendo in bar 94 of the fourth movement. Also a motif from the first movement was taken into account, which includes four sixteenth notes played with a decrescendo. The accompanying ensemble played two times with different soloists. The analysis further includes two studio recordings, in one of which the same orchestra performed but with a different soloist.

Flute solo: The *B Minor Solo for Flute and Cembalo* from the first part of the *Tafelmusik* was performed by seven flute players during the TC. The analysis focussed on several passages of two tied eighth notes played with a decrescendo. In addition, a decrescendo and crescendo within one large note, lasting seven quarter notes, was considered in the seventh and eighths bar of the first movement. In this case the decibel progression was approximated for an analysis of sub-note dynamics. Because of extensive embellishments the analysis included six performers playing the first movement.

Suites a2 and C6: Studio recordings of the first minuets of the *Suite for Orchestra in A Minor* and *C Major* (*TWV 55:a2* and *55:C6*) were analyzed regarding metrical emphases. Each minuet was repeated two times. A2 was performed by eight ensembles and C6 by four ensembles.

Experimental recordings: Ten musicians were asked to play continuous quarter notes and perform (de-)crescendi between *piano* and *forte*. The recordings were additionally used to measure the dynamic range of the instruments. Every participant performed eight crescendi and decrescendi. The experimental pieces were of two kinds so that (de-)crescendi were placed in the centre of the piece as well as on the final notes, containing one or two bars. Four musicians played the task with different instruments so that all data consisted of 16 dynamic transitions played by 14 instruments.

Dynamics are individual not only for the decision on whether a (de-)crescendo shall be performed but also for the transition length and extent of loudness. In the analysis the duration was reset when a (de-)crescendo started earlier or ended later than it was supposed to. To ensure that all transitions were continuous and distinct, data were excluded if (i) the loudness difference between the first and the last note was lower than eight decibel, (ii) the transition was not continuous but terrace-like, (iii) there were too few note events between the start and end note, or (iv) there was no transition performed. For every (de-)crescendo the decibel values and

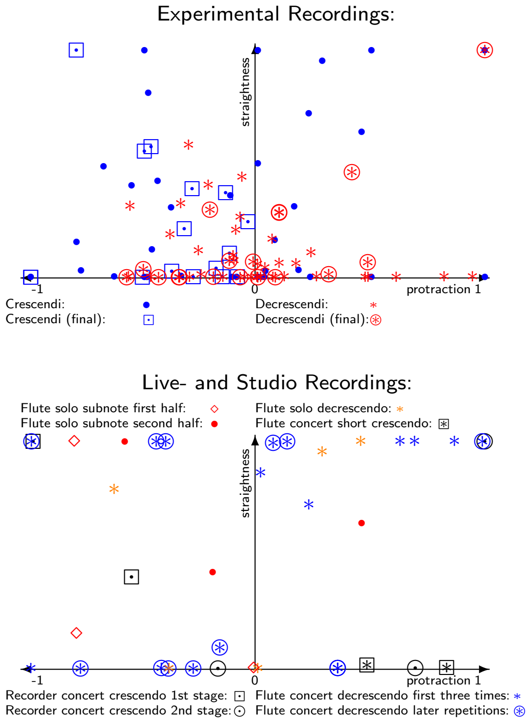

Figure 2: measured protraction and straightness parameters.

time positions were normalized. Consequently, large tempo changes would confound the curvature, so no transition was analyzed on a final ritardando or any other intense tempo changes. Afterwards, the Bézier approximation was made. Setting the first loudness value at 0 and the last at 1, the control points $0 \leq x_2 \leq 1$ and $0 \leq x_3 \leq 1$ were transformed into protraction and straightness values.

4.2 Results

All computed straightness and protraction values are plotted in Figure 2. As can be seen, both plots show no linear relationship between the two dimensions. In the experimental recordings there were no significant differences between the final and non- final position, crescendo or decrescendo. Since

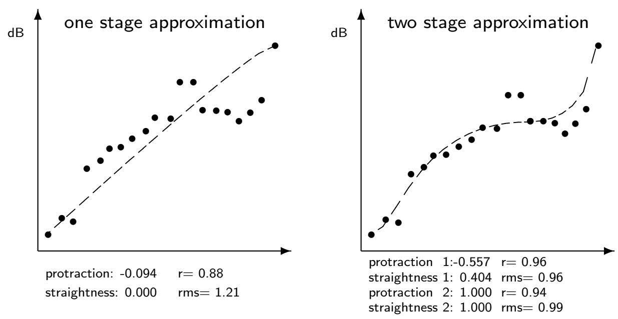

Figure 3: Crescendo characteristics in the Recorder concert. Dots: empirical data. Line: approximation.

the experimental scores only included *piano* and *forte*, all musicians played very marked (de-)crescendi. The median *p-f* range was 21.4 dB. (De-)crescendi sometimes started before the annotation and some musicians begun the crescendo with a decreased loudness. The protraction values tend to be smaller than but also near zero, whereas the straightness shows a larger distribution, particularly in the crescendi. Nevertheless, most straightness values are close to zero as well. In the live- and studio recordings many data were excluded in accordance to the preconditions in Section 4.1. For instance, from all 36 possible (de-)crescendi in the Flute solo only nine could be considered; in the Flute concert fifty percent of the 32 samples had to be excluded. From this point of view general consequences remain conjectural. Though it seems obvious that the straightness in the live- and studio recordings tends to be near the extreme poles. In both plots there is no combination of medium straightness and extreme protraction.

After the first approximation the ensemble crescendi in the Recorder concert showed a linear characteristic (not plotted in Figure 2, one instance is shown in Figure 3, left), which did not fit the empirical data well. The correlation coefficient r was rather weak and the differences of the approximation and the empirical data were large. Interestingly, the empirical data showed a flipped *S*-characteristic that contradicts the avoidance of indifferentiabilities predicted in Section 3.1. An example is plotted in Figure 3. There the crescendo consists of two stages: The first half shows a distinct increase of loudness, followed by a linear crescendo until a final jump towards the target loudness finishes it. Consequently, all crescendi were approximated anew in two steps. This increased the quality of the approximation, as the larger correlation coefficient r and the decreased root mean square value rms reflect. Shorter crescendi, for example those in the fourth movement, showed a positive protraction, which is a characteristic similar to the second stage.

The short decrescendi in the Flute concert showed diverse characteristics. At their first occurrence and in the first repetitions the protraction was greater than zero and the straightness got very large. In later repetitions the protraction fell below zero. Different characteristics were found in the Flute solo (see Figure 4). Here decrescendi were similar to the Flute concert but also linear transitions were found.

The characteristics of sub-note dynamics were taken from the Flute solo. These were unequivocally curved as assumed in Section 3.1. Figure 5 shows one example of the tied dotted

Figure 4: Decrescendo characteristics. Dots: empirical data. Line: approximation.

colors indicate normalized loudness deviation

.86 | .71 | .43 | .13 | .06 | -.21 | -.49 | -.79 | >

bar	Suite a2 1	2	3	$d\bar{B}$	bar	Suite C6 1	2	3	$d\bar{B}$
I				6.7	I				15.1
II				6.9	II				13.2
II				6.1	III				15.1
IV				6.4	IV				13.5
V				8.3	V				14.1
VI				11.2	VI				13.9
VII				9.9	VII				12.4
VIII				4.7	VIII				9.4
$\bar{\Sigma}$					$\bar{\Sigma}$				

Table 2: Loudness differences in minuet beats.

half notes that were played with a decrescendo on the first and a crescendo on the second note.

Less individual characteristics were found in the Micro Dynamics. Although the mean decibel value in the Suite C6 was larger than in a2, the proportions of loudness per measure were clearly increased for the first and decreased for the third beat. Beside the significance of the mean results, Table 2 shows the mean proportions referring to each of the eight bars in each minuet. There the cell colors reflect the loudness value referring to the distribution of all values in both tables. A typical phenomenon in Telemann's minuets is that he subdivides the eight bar phrase into two-bar sections in the first half and one bar sections in the second half. This is obviously reflected in the dynamic patterns of C6. There every second beat was accentuated as if a $\frac{3}{2}$ was being played over the $\frac{3}{4}$. Moreover, the first $\frac{3}{4}$ bar acts as pickup to the $\frac{3}{2}$, causing an ample emphasis on the first beat in all even measures. The maximum and minimum loudness differences from the mean were 2.73 dB and -2.44 dB respectively (without taking the last bar into account, for these loudness differences were independent of accents), which is twelve percent of the piano forte range of the experimental recordings. Considering that musicians play a p louder when the same piece includes a pp or even ppp, it can be assumed that the correspondent KeyOn velocity shown in Table 1 is lower for the p and larger for the f. Referring

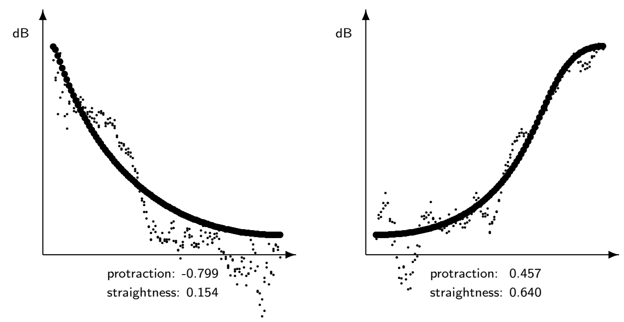

Figure 5: Approximation of sub-note dynamics in Flute solo. Small dots: empirical data. Big dots: approximation.

to a *piano-forte* difference of 55 KeyOn-velocity values the corresponding dynamic scope would be approximately 20.

4.3 Discussion

The preceding results have amply demonstrated that dynamic transitions are individually shaped; that is, any usage of mean values would not be trustworthy. Linear characteristics were found in experimental recordings. Although in the live- and studio recordings the amount of data was small and the results are therefore insignificant, the results allow to conclude that the latter characteristics are not linear. However, there are some musical facts that additionally might be plausible. The details in Figure 2 may lead to the assumption that different characteristics of crescendi and decrescendi depend on the direction of emphasis in musical structure:

Since the protraction is responsible for the early or deleted start of the core transition, it is the most influential factor. It emphasizes the musical figure, the transition itself, and the degree of expectation of the succeeding figure:

protraction < 0: The loudness transition starts early. In a crescendo the dynamics change itself becomes emphasized, whereas in a decrescendo only the very outset of the figure is loud. This results in an accentuation, as found on the later repetitions of the decrescendi in the Flute concert, or the first dotted half note decrescendo in the Flute solo.

protraction > 0: The significant part of the loudness transition starts late. This emphasizes the latter part of a crescendo by a continuous increase of the slope. This effect increases the expectation of a succeeding figure, i.e. the crescendo leads to a target. In a decrescendo the figure is present for a longer period of time and the decrescendo itself is emphasized instead.

protraction $= 0$: The transition is balanced, i.e., neutral. A possible explanation for the fact that those neutral characteristic were found in the experimental results but rarely in the live- and studio recordings might be that in the experiments there was no musical context. Some musicians complained about this, which indeed was an important condition of the task, for there should be no extraneous interference.

An increase of the straightness parameter creates a more curved characteristic. A light increase results in a more organic flow. If the protraction is not zero, the above mentioned characteristic gets emphasized. Though when the straightness value becomes very large, the area of the distinct loudness change becomes smaller and can result in a jump like transition, particularly in cases where the protraction is close to zero. The combination $straightness = 1$ and $protraction = 0$ results in quasi terraced dynamics. These are found in the later repetitions of the decrescendo in the Flute concert.

Here it bears reminding that the straightness influences the differences of the control points x_2 and x_3 of the Bézier curve. Since both control points are restricted between zero and one, the more distant the protraction value is from zero, the less is the difference between x_2 and x_3; which in turn influences the straightness. On the extreme poles -1 and 1 there is no influence on the straightness anymore. This means that if the character shall be distinctly modified by the straightness, the amount of the latter must increase when the protraction is not zero. This also explains the extreme straightness values found in the data as well as the anvil-like shape of the distributions in Figure 2.

In the experimental recordings there were some marginal dynamic transitions before the actual (de-)crescendi. A similar phenomenon was already observed by Hähnel and Berndt [5], who suggested that a slight and short preceding crescendo enables the performer to mark the actual crescendo by a clear soft start. Thus, the beginning of a dynamic transition can be stressed by a small terrace like step against the direction of the actual transition. Because this principle is, of course, admittedly speculative, further investigation is needed.

The results of the metrical accent analysis were consistent with the assumptions made at the outset of this Section. A glance at Table 2 makes it clear that the accentuation scheme follows the metrical hierarchy of the minuet. Moreover, the two bar and one bar structure in C6 is a striking example for a metrical acceleration by means of accentuation. The large differences in the last bar of a2 and C6 result from a compositional detail; the basso continuo is the only instrument that continues playing after the first beat. The whole ensemble only plays one note on the first beat in the eighths bar. The mean differences are not, however, substantially effected.

5. CONCLUSIONS

This paper developed an implementation of musical dynamics. Therefore, a distinction between the macro and micro layer of dynamics has been introduced. Furthermore, the special requirements for continuous dynamic transitions over large distances and sequences of many notes as well as within one single note were considered. These were modelled by cubic Bézier curves that allow for a flexible shaping of dynamic transitions and at the same time avoid indifferentiabilities. An analyses of human musicians proved the requirements for this large flexibility. If dynamics had been bound on musical structure, the results would have been much more alike in shape. Resemblance between different interpretations are well known for timing, but regarding dynamics the ample individual freedom found in the recordings is a striking evidence for the developed model since the analysis demonstrated that human performed dynamics can be adequately approximated by the developed functions. Ultimately, the parameterization for the different characteristics was put down to some more intuitively editable descriptors.

Acknowledgments

We like to express thanks to all musicians for their participation in the recordings and for the inspiring dialogs.

6. REFERENCES

[1] A. Berndt. Decentralizing Music, Its Performance, and Processing. In M. Schedel and D. Weymouth, editors, *Proc. of the Int. Computer Music Conf. (ICMC)*, pages 381–388, New York, USA, June 2010. International Computer Music Association, Stony Brook University.

[2] A. Berndt and T. Hähnel. Expressive Musical Timing. In *Audio Mostly 2009: 4th Conf. on Interaction with Sound—Sound and Emotion*, pages 9–16, Glasgow, Scotland, Sept. 2009. Glasgow Caledonian University, Interactive Institute/Sonic Studio Piteå.

[3] R. B. Dannenberg. The Interpretation of MIDI Velocity. In *Proc. of the Int. Computer Music Conf. (ICMC)*, pages 193–196, Tulane University, New Orleans, USA, Nov. 2006. International Computer Music Association.

[4] A. Friberg, R. Bresin, and J. Sundberg. Overview of the kth rule system for musical performance. *Advances in Cognitive Psychology, Special Issue on Music Performance*, 2(2-3):145–161, 2006.

[5] T. Hähnel and A. Berndt. Expressive Articulation for Synthetic Music Performances. In *Proc. of New Interfaces for Musical Expression (NIME) 2010*, pages 277–282, Sydney, Australia, June 2010. University of Technology Sydney.

[6] G. Mazzola, S. Göller, and S. Müller. *The Topos of Music: Geometric Logic of Concepts, Theory, and Performance*. Birkhäuser Verlag, Zurich, Switzerland, 2002.

[7] J. Meyer. *Acoustics and the Performance of Music*. Springer, 5 edition, 2009.

[8] J. Meyer. *Acoustics and the Performance of Music: Manual for Acousticians, Audio Engineers, Musicians, Architects and Musical Instrument Makers*. Modern Acoustics and Signal Processing. Springer, New York, USA, 5th edition, 2009. translator: U. Hansen.

[9] L. Mozart. *Gründliche Violinschule*. Bärenreiter, Augsburg, 1789. Faksimile- reprint (1968) of the 3rd ed.

[10] S. Pank. Der Fingersatz als ein bestimmender Faktor für Artikulation und Metrik beim Streichinstrumentenspiel. *Michaelsteiner Konferenzberichte*, 53:95–103, 1998.

[11] J. J. Quantz. *Versuch einer Anweisung die Flöte traversière zu spielen*. Bärenreiter, Berlin, 1752. Faksimile-reprint (1997).

[12] B. H. Repp. A microcosm of musical expression: Ii. quantitative analysis of pianists' dynamics in the initial measures of chopin's etude in e major. *The Journal of the Acoustical Society of America*, 105(3):1972–1988, 1999.

[13] N. P. Todd. The dynamics of dynamics: A model of musical expression. *The Journal of the Acoustical Society of America*, 91(6):3540–3550, 1992.

[14] Vienna Symphonic Library GmbH. Vienna Instruments. http://vsl.co.at/ [last visited: March 2010], 2010.

[15] G. Widmer and A. Tobudic. Playing mozart by analogy: Learning multi-level timing and dynamics strategies. *Journal of New Music Research*, 32:259–268, 2003.

[16] W. L. Windsor and E. F. Clarke. Expressive Timing and Dynamics in Real and Artificial Musical Performance: Using an Algorithm as an Analytical Tool. *Music Perception*, 15(2):127–152, 1997.