ARCHIVES OF ACOUSTICS
36, 2, 283–296 (2011)
DOI: 10.2478/v10168-011-0022-z

An Accent-Based Approach to Automatic Rendering of Piano Performance: Preliminary Auditory Evaluation[(*)]

Erica BISESI, Richard PARNCUTT

University of Graz
Center for Systematic Musicology
Merangasse 70, 8010 Graz, Austria
e-mail: {erica.bisesi, richard.parncutt}@uni-graz.at

(*received January 27, 2011; accepted March 31, 2011*)

We are exploring the relationship between accents and expression in piano performance. Accents are local events that attract a listener's attention and are either evident from the score (immanent) or added by the performer (performed). Immanent accents are associated with grouping (phrasing), metre, melody and harmony. In piano music, performed accents involve changes in timing, dynamics, articulation, and pedalling; they vary in amplitude, form, and duration. We analyzed the first eight bars of Chopin Prelude op. 28 n. 6. In a separate study, music theorists had marked grouping, melodic and harmonic accents on the score and estimated the importance (salience) of each. Here, we mathematically modeled timing and dynamics in the prelude in two ways using *Director Musices* (DM) – a software package for automatic rendering of expressive performance. The first rendering focused on phrasing following existing and tested procedures in DM. The second focused on accents – timing and dynamics in the vicinity of the accents identified by the theorists. In an informal listening test, 10 out of 12 participants (5 of 6 musicians and 5 of 6 non-musicians) preferred the accent-based formulation, and several stated that it had more variation of timing and dynamics from one phrase to the next.

Keywords: piano performance, expression, accents, musical analysis, Director Musices.

Notations

Abbreviations

- DM – Director Musices,
- P – phrasing-based formulation rendering,
- A – accent-based formulation rendering,
- M – musicians,
- NM – non-musicians.

[(*)] This paper was presented during the ViennaTalk 2010 conference (Bisesi, Parncutt, 2010).

1. Introduction

In a broad definition of accent (Parncutt, 2003), the main accents types are as follows. Grouping accents occur at starts and ends of note groups at different hierarchical levels. Metrical accents are similarly hierarchical and relate to the underlying beat (e.g. Wojcik, Kostek, 2008; 2010). Melodic accents may be turns (i.e., peaks and valleys of the melodic contour), and skips (i.e. disjoint intervals between consecutive tones). The wider the interval before a tone, the stronger its accent, and rising skips produce stronger accents than falling. Harmonic accents correspond to events associated to harmonic tension, and occur at harmonic changes and at harmonic dissonances. Dynamic accents are those explicitly marked in the score. Timbral accents correspond to change of instrument and articulatory accents correspond to legato and staccato. According to Lerdahl and Jackendoff (1983), musical accents may be classified as *structural* – those assumed to be intrinsic to the notated score – and *expressive*, which are the ones added to the score by a performer. Parncutt (2003) labeled these two categories *immanent* and *performed*. Both have aspects associated with the four primary perceptual attributes of a sound: time, pitch, loudness and timbre. Musical experience suggests that immanent accents may define categories of perceptual attributes whose boundaries are determined perceptually and within which performance parameters can be manipulated. Table 1 shows how immanent and performed accents can be classified according to perceptual attribute.

Table 1. Parncutt's (2003) taxonomy of musical accents.

Accents	Immanent	Performed
time	grouping	agogic (onset time)
	metrical	articulatory (duration)
pitch	melodic	intonation
	harmonic	
loudness	dynamic	stress
timbre	instrument	coloration
	orchestration	

The relationship between immanent and performed accents is that performers tend to "bring out" immanent accents, i.e. to attract the listener's attention to them. For example, a performer may slow the tempo or add extra time in the vicinity of certain kinds of immanent accent, or change dynamics or articulation (the degree of staccato (separation) or legato (overlap) of successive events) in consistent ways. This relationship is complex and depends on many factors such as musical and personal style, local and cultural context, intended emotion or meaning, and acoustical and technical constraints.

Our approach to music analysis and interpretation assumes that musical events that attract attention (accents) can function as a vehicle for emotional expression. So there is an interesting open question about the relationship between emotion and accent – both immanent emotion (SLOBODA, 1991) and performed emotion (the topic of our research). Sloboda identified simple structures like sequences (related to structures of grouping accents), appoggiaturas (harmonic accents or – according to *Director Musices* – *melodic charge*), and new or unexpected harmonies (harmonic accents or *harmonic charge*) which are associated with emotional responses. We can link Sloboda's main findings with Parncutt theory of accents, as follows: (1) a harmonic descending cycle of fifths to tonic is an event that arouses expectation of tonic arrival; (2) melodic appoggiaturas can be interpreted as harmonic accents (Sundberg: *melodic charge*); (3) melodic or harmonic sequences involve structural accents (at the start of each repetition, at a different pitch); (4) enharmonic changes involves harmonic accents (sudden change of harmony or tonality); (5) harmonic or melodic acceleration to cadence corresponds to an increase in temporal density of (harmonic or melodic) accents (number of accents per unit time), implying that emotional intensity depends on the salience and temporal density of accents; (6) delay of a final cadence may be regarded as an unexpected structural accent; (7) a new or unprepared harmony corresponds to a harmonic accent; (8) a sudden dynamic or textural change may be related to any accent (something that attracts attention); (9) repeated syncopation is a change in metrical accent pattern; (10) a prominent event that came earlier than prepared for is also something that attracts attention. Whereas Sloboda focused on what we call immanent accents (emotional expressivity that is somehow inherent in the score), we will now address the relation between immanent and performed accents – with a common underlying concept of accentuation as psychological salience (perceptual importance or probability of noticing) (PARNCUTT, 1989).

2. An accent-based approach to automatic rendering of expressive piano performance

The following stages have been planned for our investigation of the relationship between immanent accents and general expressive features (performed accents) in piano music: (i) the empirical determination of immanent accent locations in the musical score and their relative salience (musical analysis); (ii) the setting of expressive parameters (timing and dynamics) for each performed accent (a style-dependent mathematical and physical modeling); (iii) the numerical implementation of the mathematical model into a new formulation of *Director Musices* (DM) (Director Musices, 2007) – a software package for automatic rendering of expressive performance (computational modeling).

2.1. Music analysis

Notes and groups of notes do not divide easily into two categories, accented and unaccented. The degree of accentuation varies on a continuous scale. Here we use the term *salience* for the importance of a note. The salience of an immanent accent may be considered to be its perceptual importance when the music is heard in a typical expressive performance, or even in a deadpan performance.

Figure 1 shows a general example of accentuation (the one preferred by most of music theorists). Because we are talking about accentuation that arises only from the musical structure and not from performance, we may use the term *immanent accentuation.* Here, immanent accents of the first eight bars of Chopin Prelude op. 28 n. 6 are divided into three types: phrasing (or serial grouping), melodic (or contour), and harmonic (or dissonance). Accents are subjectively assigned a salience level ranging from 1 to 5. This is indicated by the size of the squares at melodic accents (C)[1] and harmonic accents (H).

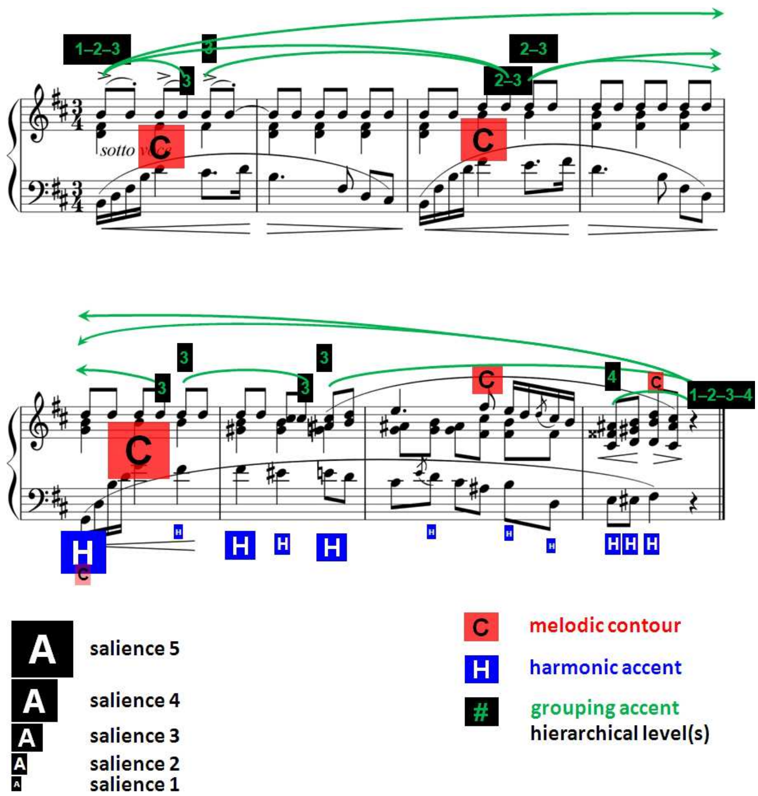

Fig. 1. Example of subjective analysis of accents in the first eight bars of Chopin Prelude op. 28 n. 6. Hierarchical phrasing is indicated by curved lines from the start and the end of each phrase, subphrase and sub-subphrase. Numbers in the boxes refer respectively to the start (boxes over the line) and the end (boxes below the line) of the main phrase (1), of subphrases (2) and sub-subphrases (3 and 4). Melodic contours and harmonic accents are indicated by boxes of different size, depending on their salience (levels from 1 to 5).

[1] We use the letter C for melodic accents because they refer to the contour of the melody and the letter M may be used for metrical accents.

To mark the phrasing accents on a score, it is necessary first to describe the piece's hierarchical phrasing structure. In this example, music theorists first regarded the entire excerpt as one long phrase (indicated by hierarchical level 1 in the boxes), then divided this into two subphrases (hierarchical level 2) of nominally equal importance. Then they divided subphrases into sub-subphrases (hierarchical level 3 and 4). Figure 1 shows the boundaries of each phrase, subphrase and sub-subphrase denoted by a curved line from its beginning (indicated by numbered boxes over the line) and its end (numbered boxes below the line). The starts and ends of subphrases and sub-subphrases are marked intuitively according to repetitions of motives (as in the first part of the excerpt) or introduction of new structural elements (as in the second part of the excerpt).

To identify the melodic accents, music theorists have labeled the highest and lowest tones of the whole melody, then labeled the local peaks and valleys, i.e. the highest and lowest pitches in a given phrase. The third melodic accent (C) corresponds to a valley; all the other are melodic peaks. The salience of the melodic peak in bar 5 is bigger than the salience of the first two peaks, not only because is it higher in pitch, but also because it occurs at the third repetition of the same motive. As peaks normally have more salience than valleys, the melodic valley at the beginning of the fifth bar has low salience (of level 2). The last two melodic peaks in bars 7 and 8 are less salient than the others, if we assume that the melody in the right hand in the last sub-subphrase is less important than the left-hand melody in previous bars; further, their relative salience is different, because the pitch of the first one is higher than the pitch of the second one.

The harmonic accent of a chord in a chord progression has several components: roughness, harmonic ambiguity, harmonic relationship to context, and familiarity or expectedness. These components can change independently, so we need to consider all of them. The first chord in bar 5 is an unexpected new chord, so we have marked it as a harmonic accent of salience 4. The first accent in bar 6 is regarded as a dissonance carrying more tension than its resolution (on the following accent). Similarly, the new chord at the beginning of the last sub-subphrase of level 3 is interesting, because it carries out a chromatic modulation from F sharp minor to B minor.

2.2. Mathematical modeling

The formulation of a physical model of expressive performance involves adjustments in tempo and dynamics curves by means of a set of prescriptions, enabling different interpretations of a musical score. The current version of DM operates on variables such as inter-onset duration, amplitude, and pitch, modeling aspects of structure such as phrasing, articulation and intonation. In our adapted version of this software, we adjusted tempo and dynamics in the vicinity of specific accents. Our aim is to produce computer-generated performances that sound "musical" and "natural" by relating expressive features of a performance not only to structural properties, but also by accounting for local events (individual notes corresponding to accents) in a systematic way.

Many studies have suggested a relationship between musical motion and physical movement (Clynes, 1977; Friberg, Sundberg, 1999; Repp, 1992; Todd, 1992; 1995). For instance, Todd compared the variation of tempo in music with velocity in the equations of elementary mechanics (Todd, 1995). This mathematics can be incorporated into a model of performed accent that predicts timing and/or dynamics in the vicinity of an accent, where the strength of the accent corresponds to the height of the peak, the width of the curve accounts for the duration of the event, and the graphic shape is representative of finer expressive elements. For instance, melodic accent strength may depend on the size of the interval preceding the corresponding tone (in skips), the shape of the melodic contour and the distance from the mean pitch of the local context (in turns). The model should be asymmetrical with respect to peaks and valleys (Parncutt, 2003). Figure 2 shows some examples of mathematical functions that can be used to model some of these effects. As all the functions are given in a parametric representation, the shape of the timing and/or dynamics curves in the vicinity of an accent may be determined by the kind of the function involved, its width and peak amplitude (the area below the curve). The change of curvature (second derivative) and asymmetry of the curves in tempo and dynamic level are particularly interesting from a musical point of view, as they can explain a range of expressive variations. For instance, a salient event in the score may be emphasized in performance by temporal preparation and addition of elements of surprise – that are suitable for harmonic changes and dissonances, while a monotonic increase (or decrease) in tempo and/or dynamics may sound more natural and appropriate for less expressive gestures. The degree of asymmetry in the global event curve may depend on the salience of a particular musical attribute of the accent.

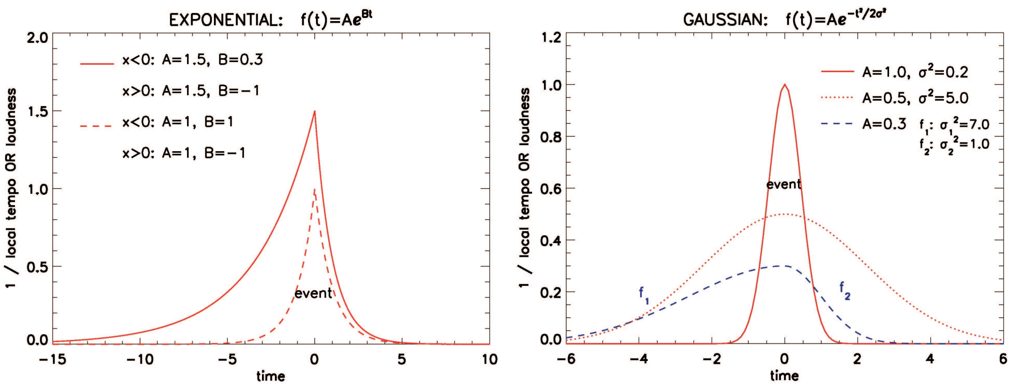

Fig. 2. Examples of analytical functions to model gradual changes in tempo and dynamic level in the vicinity of accents.

2.3. Computational modeling

Like other areas of research, modern musicology can take advantage of sophisticated computer models in diverse areas, including the analysis of scores and sound files (music information retrieval), composition, and expressive performance. But despite significant progress in music performance analysis and rendering in recent years, computer generated performances are still clearly artistically inferior to real ones. So far, generative algorithms cannot reliably render musically convincing performances in a variety of styles, nor has a single theoretical approach begun to dominate the field. The modeling of performance expression may be considered one of today's most important unsolved problems in music psychology, suggesting that considerable progress is still possible in this area.

Given that background, our project investigates the complex web of relationships that link musical structure to musical expression:

- How and to what extent are timing and dynamics related to musical accents (in a broad definition) in expressive piano music?
- Which computational models best account for changes in timing and dynamics in a vicinity of accents?
- How can such models best be incorporated into existing computing environments for simulating music expression?

2.3.1. Director Musices

DM (Director Musices, 2007) was developed in the 1980s and 1990s by Anders Friberg, Lars Frydén and Johan Sundberg in a long-term research project at the Royal Institute of Technology, Stockholm (Friberg *et al.*, 2006; Sundberg, 1988). It comprises performance "rules": mathematically defined conventions of music performance that change specific note properties such as duration and intensity. By manipulating program parameters, meta-performers can change the degree and kind of expression by adjusting the extent to which each rule is applied. The program is implemented in Common Lisp and in its current formulation is available free as a stand-alone application for Macintosh or Windows. The main advantages of DM for our purposes are that it already works well in its current form and the architecture and code are flexible enough to permit gradual evolution of both structure and mathematical formulations. For these reasons it is an ideal environment for testing a new theory of expression or developing educational applications.

Several of DM's rules can be interpreted in terms of Parncutt's (2003) taxonomy of accents. For example, DM invokes specific changes of timing and dynamics in the temporal vicinity of the peak of a melodic contour of a phrase. In Parncutt's model, a melodic contour peak is regarded as a melodic accent, and any accent can or should be emphasized by a slowing of tempo in the vicinity of the accent (or by the insertion of short time delays) and/or by a temporary increase (or decrease) of dynamic level. Comparisons of this kind suggest that a conflation of the two models may yield new insights into expressive performance and possibly lead to artistically superior computer-rendered performances. As a first attempt, we can incorporate aspects of Parncutt's accent model into the code, retaining existing formulations that correspond to accents, so that existing procedures in the model can be reinterpreted in relation to accent theory. Specifically, we can adjust local tempo and dynamic curves in the vicinity of a specific accent. Moreover, we can implement Parncutt's model within DM by adding rules that relate expressive features of a performance not only to structural properties (i.e. different levels of phrasing), but also to individual events (notes and chords) in a systematic way.

An example of expressive rendering with the old and new formulation is discussed in next paragraph.

3. Methods

In the first stage, we created two different renditions of the first eight bars of Chopin Prelude op. 28 n. 6 according to the interpretation outlined by the analysis of Fig. 1: the first based on the current version of DM, the second based on a new version of that code. Second, we compared performance renderings by mean of an informal listening test.

3.1. Preparation of sound examples

Musical expression is mainly based on musical structure (Clarke, Baker–Short, 1987), and the most important aspect structure for expression in mainstream classical Western music is phrasing (Sundberg *et al.*, 2003; Todd, 1985). Once the phrase boundaries are marked in the score, the DM's Phrase-Arc rule links each phrase to arch-like tempo and sound-level curves according to a given parameterization (eight free parameters for each hierarchical level). The parameters determine the shape of the profiles in timing and dynamics, the turning positions (peaks) of the phrases, scaling factors for inter-onset-level and sound level, scaling factors for next phrase level ending and for second next phrase level ending, scaling factors for acceleration in the beginning and for last note (P, phrasing-based formulation) (Director Musices, 2007). Figure 3 shows an example of application of this rule to the first eight bars of Prelude op. 28 n. 6 as analyzed by music theorists (Fig. 1): the solid (light) lines in the upper panels show the duration relative to the nominal duration of each note as a function of its position in the score, and the solid (light) line in the lower panel corresponds to the difference in sound level from the default value as a function of the note position. Curves in the left and right upper panels respectively refer to the duration differences in the two thematic voices of the piece, i.e. the bass (B) and the

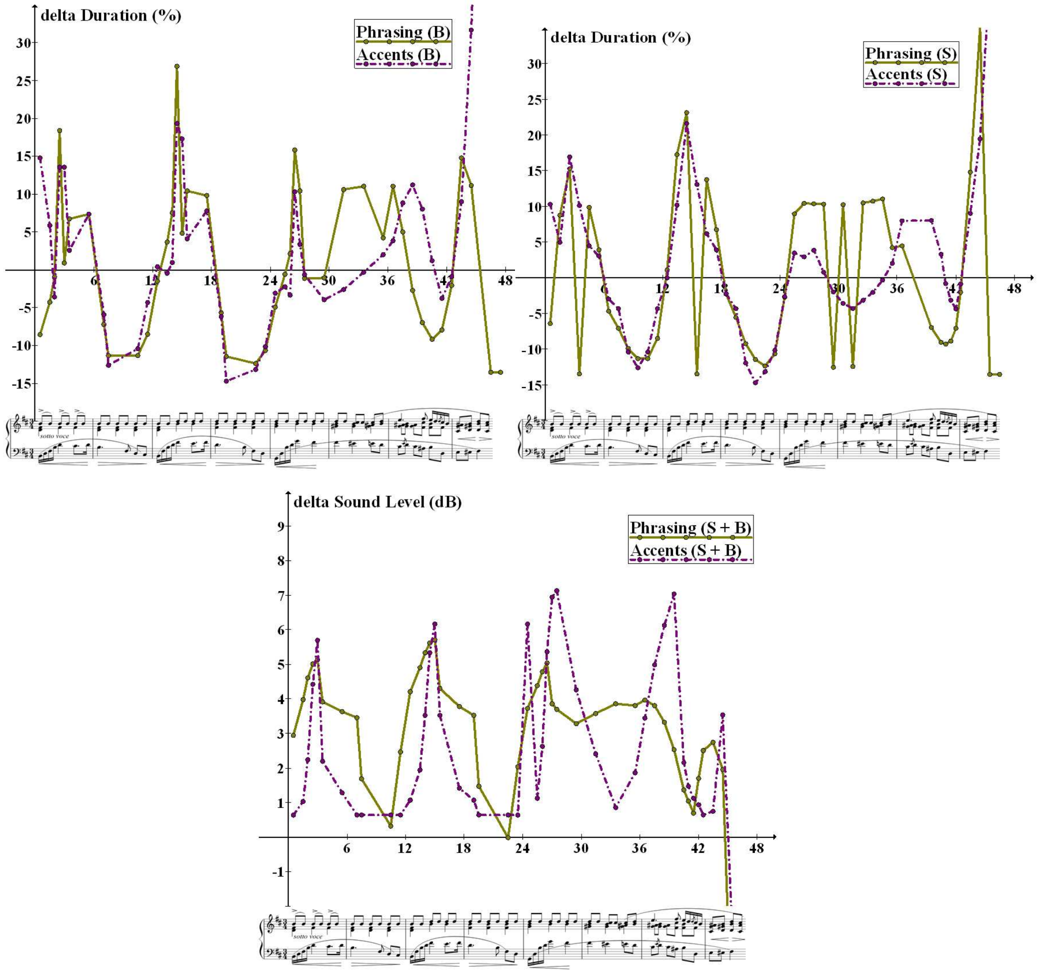

Fig. 3. Example of mathematical modeling of timing (upper panels) and dynamics (lower panel) based on an analysis of the Chopin Prelude op. 28 n. 6 in Fig. 1. The upper plots show the duration relative to the nominal duration of each note as a function of its position in the score, respectively for the bass (B, upper left panel) and the soprano (S, upper right panel); curves in the lower plot correspond to the superposition of differences in sound level from the default values for bass and soprano, as a function of the note position. Solid (light) lines refer to the phrasing-based formulation – the current formulation of DM, which is dominated by phrasing. Dashed-dotted (dark) lines refer to the accent-based formulation – our new formulation of DM, based on accents.

soprano (S); the curve in the lower panel is the superimposed difference in sound level of bass and soprano. We choose to represent only these two voices, because they are the only ones where we have applied accents (see Fig. 1), so comparison with default values of timing and dynamics in the accent-based formulation (as discussed below), and therefore between the two formulations, makes sense. In this example, phrasing has been modeled with the Phrase-Arc rule and we have set parameter values subjectively by trial and error. In DM, all rules have a quantity parameter k, which controls the general effect. A k value of 1 corresponds approximately to a normal (default) application of the rule (Director Musices, 2007). This is, however, dependent on the musical context. In this example, Phrase-Arc quantities (magnitude of variations in timing and dynamics) at the different phrasing levels are as follows: 0.1 at phrasing level 1 (the longest phrase); 0.7 at level 2; 1.8 at level 3; and 0.4 at level 4. As phrases in Fig. 1 often begin with an ascending leap, to improve phrasing we also used Leap-Tone-Duration (this rule shortens the first note of an ascending leap and lengthens the first note of a descending leap).

The dashed-dotted (dark) lines in Fig. 3 refer to the accent-based formulation (A) – our new formulation of DM. Here, expression at each accent marked in Fig. 1 is modeled by means of two new functions, one for timing (duration) and one for dynamics (sound level) in the vicinity of the accent (respectively, Accent-Main-Dr and Accent-Main-Sl). Each function admits five free parameters: the event peak, the width of the interval preceding the accent, the width of the interval following the accent, the shape of the curve before the peak, and the shape of the curve after the peak. In this case, accents are modeled with exponentials and Gaussians of different peaks and widths, corresponding to accent saliences in the analysis of Fig. 1. As in the previous formulation of DM, all the new rules are additive: when a tone or a chord has more than an accent, profiles in timing and dynamics account for the global (superimposed) effect of all the accents (a linear combination). As in the previous case, we chose the best settings of new rules, according to our preliminary subjective auditory evaluation. The only rules from the previous formulation of DM that are applied here are Leap-Tone-Duration (with the same quantity as in the phrasing-based model) and Final-Ritard (quantity $k = 2.5$). These two additional rules are applied to the same extent in the two formulations, which makes them comparable. Anyway, the effect of Leap-Tone-Duration and Final-Ritard on the renditions is minimal; therefore the two formulations differ essentially in the use of Phrase-Arc (in the phrasing-based model) and Accent-Main-Dr and Accent-Main-Sl (in the accent-based model).

In both formulations, the leading strategy has been to emphasize phrase boundaries and accents as better as possible (according to the analysis of Fig. 1), to find out the combination of free parameters corresponding to models the more similar each other (in order to outline their intrinsic differences), and to produce the best possible sound files (according to our subjective auditory evaluation).

3.2. Preliminary auditory evaluation

Differences and similarities in the two formulations discussed above were further investigated in an informal perceptual test. In this stage, sound files rendered by the two DM' formulations were submitted to 12 participants for an auditory evaluation based on an exploratory questionnaire.

3.2.1. Participants

Participants were 6 musicians and 6 non-musicians. Musicians were pianists with an academic degree in piano performance or people with experience in public performance. As non-musicians, we interviewed amateur musicians (including choristers and a dance teacher) and music lovers (people who frequently attend concerts and listen to CDs of classical music).

3.2.2. Equipment

Auditory stimuli were outputs of the two mathematical formulations. In order to improve the quality of the renditions, original MIDI files produced by DM were converted into WAV format by mean of the commercial software Music Creator 5.0.4.23.

3.2.3. Procedure

All participants heard the sequence of two sound files two times. The order of the two files was randomized. They were asked to judge which one sounded better, and to provide a qualitative motivation for their choice. All interviews have been made in participants' native language (German, Italian, and Spanish), and then transcribed in English.

4. Results

10 out of 12 participants (83.3%) preferred the sound file based on the accent-based formulation. The 2 participants who preferred the sound file based on the phrasing-based formulation were a musician (with a Bachelor degree in Piano Performance) and a music lover. Musicians' descriptions of the two interpretations tended to be more analytic, while non-musicians referred mainly to emotions and free associations. All participants agreed with that the two interpretations are very similar, so the task was quite difficult. Participants who preferred the interpretation based on accents used words like "more expressive", "more natural", "more flowing", "more catchy", "smoother", "softer", "less foregone", "it contains more details", "it expresses more music", "one can distinguish among the different voices", "it excites me more", "it looks more played by an human", "it is warmer", "more elastic", "brighter", "more airy", "a mouthful of air", "a feeling of openness", "it gives a sense of movement". To give some examples, musicians' opinion were (words in square brackets are authors' comments): "The first [A, i.e. accent-based formulation] is more expressive", "Difficult to judge, they sound very similar. I prefer the first [A], because dynamics and tempo are independent. The climaxes are more evident in the dynamics than in the tempo. The first is more natural, more flowing", "The second [A] seems to me to be played more

flowingly. The first [P, i.e. phrasing-based formulation] limps a bit". On the other side, motivations provided by some non-musicians are: "I chose the first [A] because it is more catchy and smoother. The second [P] is harder", "I chose the first [A] because it is softer. The second [P] is more resolute, more marked. What is disturbing me there is that the upper voice is played as loud as the lower voice, I cannot hear any difference", "Both are well played, but the second [A] excites me more. This is my first impression, but also after a second listening I preserve this opinion. I can hear more details here, and recognize that the music is played by a human. The first [P] seems to be produced by a machine, so it is less exiting. Maybe the first is a real performance, and the second some kind of manipulation of that performance. The difference is more evident after first bars: in the first 8–10 seconds [first 2 bars] the two look to be indistinguishable". Two participants (16.7%) preferred the interpretation based on phrasing. One of them judged the little breath before the highest notes in the lower voice as crucial, because it gives more sense to the phrases. We find this observation very interesting. This effect is due to the Leap-Tone-Duration rule that is conceived in the framework of the phrasing model, and is therefore more effective in that case. We will consider this point in future improvements of our model. Another participant preferred the phrasing-based interpretation because one could "hear more sound there". We interpret this observation as an effect of the sound level redistribution among voices in the phrasing-based formulation: as Phrase-Arc acts contemporary on timing and dynamics, this rule is applied to different voices in the same way in order to synchronize them. This effect is consistent with other participants' capability of distinguishing the voices in the accent-based formulation, where the model allows working out the voices independently one another. 33.3% of participants noticed that in the accent-based formulation rendering the subphrases were more variable in their timing and dynamics.

5. Discussion and conclusions

Figure 3 suggests that the accent procedure can closely approximate the patterns of timing and dynamics obtained in the phrasing-based formulation, without the need for any other principle. This is evidence for an intrinsic relationship between phrase structure and accents in the first part of the piece. The opposite does not occur: DM's phrasing algorithms (the only Phrase-Arc rule) cannot reproduce all timing and dynamic fluctuations at the local note level, as it is evident in the second part of the excerpt. The expressive rendering based on accent theory sounds promising. It considers local events as pillars supporting a phrase or subphrase (bottom-up approach), and not a subtler realization of the phrasing structure (top-down approach). In its previous formulation, DM has a mixture of top-down and bottom-up components (local and global rules); our simulations suggest that in the new formulation the bottom-up approach can work alone, which would make the model more parsimonious. Another advantage of the accent-based formulation is that different subphrases can be modeled independently, leading to higher variability in the profiles of timing and dynamics and hence a wider spectrum of performances.

We tested the two models by mean of a preliminary auditory evaluation, carried out with two groups of participants (6 musicians and 6 non-musicians). In both groups, 5 out of 6 of participants qualitatively judged the interpretation based on accents to be the best rendering of the piece, and 2 out of 6 spontaneously noticed and commented on the higher local variability of the accent-based formulation rendering.

Although the selected musical piece has a slow character (i.e. it is usually played at slow tempo), we reasonably expect that the above conclusions will be valid also for fast character pieces. Actually, the main difference between the two modeling approaches to automatic rendering with DM is the higher account of local variability in the accent-based formulation, and therefore the higher emphasis on the intrinsic relationship between phrase structure and accents. This is also the main reason of preference of the accent-based formulation rendering by most of the participants. As these aspects (phrasing, accents) are intrinsic to the structure of the piece and do not depend on the piece character, we expect that the subjective acceptance of the two examined rendering approaches will be affected neither by the rendering tempo (slow or fast), nor by mood markings (i.e. *allegro*, *grave* or *cantabile*).

These preliminary findings are encouraging and motivate us to continue to develop the model in this new direction.

Acknowledgments

This work was supported by the Lise Meitner Project M 1186-N23 ("Measuring and modeling expression in piano performance"), sponsored by the Austrian *Fonds zur Förderung der wissenschaftlichen Forschung (FWF)*. We thank Anders Friberg (KTH, Stockholm) for valuable advice and support with the implementation of new rules in DM.

References

1. Bisesi E., Parncutt R. (2010), *An accent-based approach to automatic rendering of piano performance*, Proceedings of the Second Vienna Talk on Music Acoustics, pp. 26–30, Vienna.
2. Clarke E., Baker–Short C. (1987), *The imitation of perceived rubato: A preliminary study*, Psychology of Music, **15**, 58–75.
3. Clynes M. (1977), *Sentics: The touch of emotions*, Doubleday Anchor, New York.

4. Director Musices Download Page (2007), Retrieved August 14th, 2007, from KTH, Department of Speech, Music and Hearing:
http://www.speech.kth.se/music/performance/download/dm-download.html

5. Friberg A., Sundberg J. (1999), *Does music performance allude to locomotion? A model of final ritardandi derived from measurements of stopping runners*, Journal of the Acoustical Society of America, **105**, 3, 1469–1483.

6. Friberg A., Bresin R., Sundberg J. (2006), *Overview of the KTH rule system for musical performance*, Advances in Cognitive Psychology, **2**, 3, 145–161.

7. Lerdahl F., Jackendoff R. (1983), *A generative theory of tonal music*, MIT Press, Cambridge, MA.

8. Parncutt R. (1989), *Harmony: A psychoacoustical approach*, Springer Verlag, Berlin.

9. Parncutt R. (2003), *Accents and expression in piano performance*, [in:] *Perspektiven und Methoden einer Systemischen Musikwissenschaft (Festschrift Fricke)*, Niemöller K.V. [Ed.], pp. 163–185, Peter Lang, Frankfurt/Main, Germany.

10. Repp B.H. (1992), *Diversity and commonality in music performance: An analysis of timing microstructure in Schumann's Träumerei*, Journal of the Acoustical Society of America, **92**, 2546–2568.

11. Sloboda J.A. (1991), *Music structure and emotional response: Some empirical findings*, Psychology of Music, **19**, 110–120.

12. Sundberg J. (1988), *Computer synthesis of music performance*, [in:] *Generative Processes in Music*, Sloboda J.A. [Ed.], pp. 52–70, The Clarendon Press, Oxford.

13. Sundberg J., Friberg A., Bresin R. (2003), *Attempts to reproduce a pianist's expressive timing with Director Musices performance rules*, Journal of New Music Research, **32**, 3, 317–325.

14. Todd N.P. (1985), *A model of expressive timing in tonal music*, Music Perception, **3**, 33–58.

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

16. Todd N.P. (1995), *The kinematics of musical expression*, Journal of the Acoustical Society of America, **97**, 1940–1949.

17. Wojcik J., Kostek B. (2008), *Computational complexity of the algorithm creating hypermetric rhythmic hypotheses*, Archives of Acoustics, **33**, 1, 57–63.

18. Wojcik J., Kostek B. (2010), *Representations of music in ranking rhythmic hypotheses*, [in:] *Advances in Music Information Retrieval*, Ras Z.W., Wieczorkowska A. [Eds.], SCI 274, pp. 39–64, Springer Verlag, Berlin Heidelberg.