AN ACCENT-BASED APPROACH TO AUTOMATIC RENDERING OF PIANO PERFORMANCE

Erica Bisesi, and Richard Parncutt

Center for Systematic Musicology
Karl-Franzens-Universität Graz, Austria
erica.bisesi@uni-graz.at

ABSTRACT

We are exploring the complex 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 (temporal, serial) grouping (phrasing), metre (downbeats), melody (peaks, leaps) and harmony (or dissonance). In piano music, performed accents involve changes in timing, dynamics, articulation, and pedalling; they vary in amplitude, form (amplitude as a function of time), and duration (the period of time during which the timing or dynamics are affected). We are analyzing a selection of Chopin Preludes using a novel method that combines aspects of the generative approach of Lerdahl and Jackendoff (1983), the modeling approach of Sundberg (1988) and the accent-based approach of Parncutt (2003). In the first stage, pianists and music theorists mark grouping, melodic and harmonic accents on the score, estimate the importance (salience) of each, and discuss their interrelationships. In the second stage, we mathematically model timing and dynamics in the vicinity of selected accents using an extended version of *Director Musices* – a software package for automatic rendering of expressive performance.

1. MUSICAL ACCENTS AS EXPRESSIVE ELEMENTS

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. 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 of the notated score – and *expressive*, which are the ones added to the score by a performer. Parncutt (2003) labelled 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.

ACCENTS	*IMMANENT*	*PERFORMED*
time	grouping metrical	agogic (onset time) articulatory (duration)
pitch	melodic harmonic	intonation
loudness	dynamic	stress
timbre	instrument orchestration	coloration

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

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 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* (DM) – "melodic charge"), and new or unexpected harmonies (harmonic accents or "harmonic charge") which are associated with emotional responses. As there exists an interesting relationship between Parncutt's (2003) model of accents and data of Sloboda (1991), 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 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 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 PIANO PERFORMANCE

The following stages are 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 (style-dependent mathematical and physical modelling).

2.1. Musical 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" to describe the importance of a note. The salience of an immanent accent may be considered to be the same as perceptual importance when the music is heard in a typical expressive performance, or even in a deadpan performance.

Fig. 1 shows a general example of accentuation. 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 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)**.

To identify the melodic accents, we 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 it is 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 of one another, 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 is interesting, because it carries out a chromatic modulation from F sharp minor to B minor.

To mark the phrasing accents on a score, it is necessary first to describe the piece's hierarchical phrasing structure. In this example, we first regard the entire excerpt as one long phrase (indicated by hierarchical level **1** in the boxes), then divide this into two subphrases (hierarchical level **2**) of nominally equal importance. Then we divide each subphrase into two sub-subphrases (hierarchical level **3**). Fig. 1 shows the boundaries of each phrase, subphrase and sub-subphrase 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 are marked intuitively by dividing the main phrase into excerpts of equal length, while sub-subphrases involve repetitions of motives (as in first two sub-subphrases) or introduction of new structural elements (as in the last sub-subphrase).

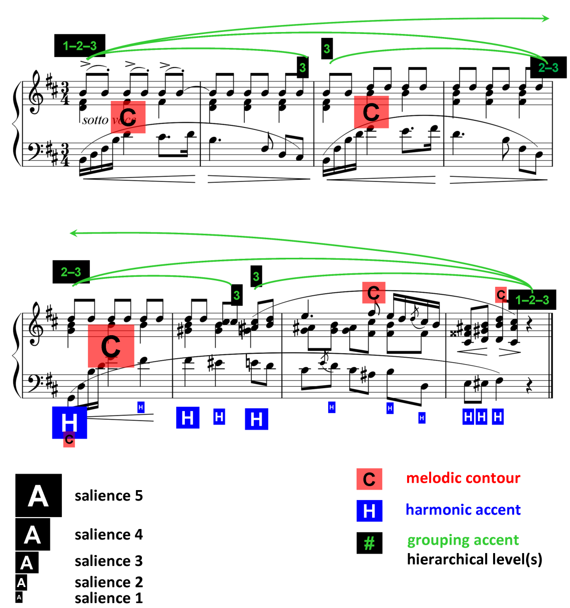

Figure 1: Subjective analysis of accents in the first eight bars of Chopin Prelude op. 28 n. 6. Melodic contours and harmonic accents are indicated by boxes of different size, depending on their salience (levels from 1 to 5). Hierarchical phrasing is indicated by curved lines from the start and the end 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).

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 tone, inter-onset duration, amplitude, and pitch, modeling global aspects of structure such as phrasing, articulation and intonation. In our adapted version of this software, we adjust local tempo and dynamics curves in the vicinity of a specific accent. We expect that implementation of Parncutt's

[1] We use the letter C for melodic accents because they only refer to contour.

model into DM will improve the musical qualities of its renditions, by relating expressive features of a performance not only to global or intermediate 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 and 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).

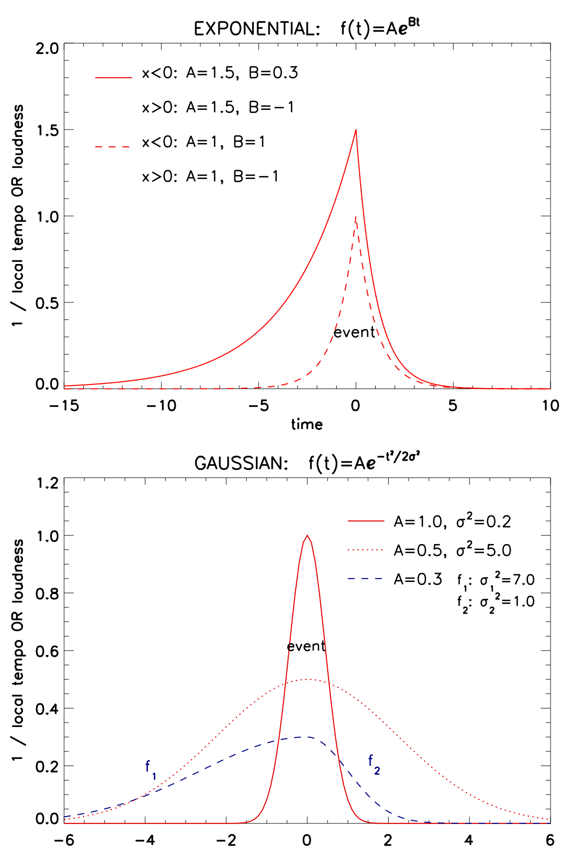

Figure 2: Examples of analytical functions to model gradual changes in tempo and dynamic level in the vicinity of accents.

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 will sound more natural and appropriate for less expressive gestures. The degree of asymmetry in the global event curve may reflect the salience of a particular musical attribute of the accent.

3. COMPUTATIONAL MODELING OF MUSICAL EXPRESSION

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. 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 is one of today's most important unsolved problems in music psychology. Considerable progress is still possible in the automatic generation of musically satisfying or acceptable renderings. Given that background, out project will investigate 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?

We will address these questions by incorporating the accent theory of Parncutt (2003) into Director Musices – a performance rendering system that introduces expressive deviations into input score files.

3.1. Director Musices

Director Musices (*http://www.speech.kth.se/music/performance /download/dm-download.html*) 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, Bresin, and Sundberg, 2006; Sundberg, 1988). It comprises performance rules (mathematically defined conventions of music performance) that change specific note properties, including timing, 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 (or all rules are) applied. The program is implemented in Common Lisp and in its current formulation is available free as a stand-alone application both for Macintosh and Windows platforms. The main advantages of DM for our purposes are *(i)* it already works well in its current form; *(ii)* the architecture and code are flexible enough to permit gradual evolution of both structure and mathematical formulations; *(iii)* it is an ideal environment to develop specific tools for perceptual experiments and/or educational applications.

Several of the rules presented in DM can be interpreted in terms of Parncutt's taxonomy of accents. For example, DM invokes specific changes of timing and dynamics in the temporal vicinity of the peak of a melodic contour. In Parncutt's (2003) 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 will adjust local tempo and dynamic curves in the vicinity of a specific accent. Moreover, we will improve DM' potentialities by implementing Parncutt's model into new rules, relating expressive features of a performance not only to global or intermediate structural properties (i.e. different levels of phrasing), but also accounting for local events (individual notes corresponding to accents) in a systematic way.

An example of expressive rendering with the two formulations – respectively with the old and the new systems of rules – is discussed in next paragraph.

3.2. Preparation of sound examples

A major principle in the current formulation of DM is *phrasing* – a prominent criterion in the majority of music performances (Sundberg, Friberg & Bresin, 2003). Once the phrase boundaries are marked in the score, the Phrase-Arc rule associates to each phrase an arch-like tempo curve and a sound level according to a given parametrization. Among the others, parameters involve the shape of the profiles in timing and dynamics, scaling factors for inter-onset-level and sound level, and turning position in the phrase. With the current rule system, DM can model also some aspects of tonal tension: the Melodic Charge rule emphasizes tones that are far away from the current root of the chord on the circle of fifths, and the Harmonic Charge rule emphasizes chords that are far away from the current key on the circle of fifths. Fig. 3 shows an example of application of these rules to the score of Fig. 1: solid (blue) line in the upper plot corresponds to the duration difference from nominal duration expressed in percent of each tone's duration as a function of the note position in the score, and solid (blue) line in the lower plot corresponds to the difference in sound level from the default value as a function of the note position. In this example, phrasing is modeled with the best settings of Phrase-Arc, according to a preliminary auditory evaluation based on our ears. Here, levels of phrasing 1 and 2 have quantity 0.1 and level 3 has quantity 0.6. To improve phrasing and articulation, we also used Phrase-Articulation (introduction of micropauses after phrase and subphrase boundaries, and lengthening of the last note in phrases), Score-Legato-Art (tone overlaps for notated legato), Leap-Tone-Duration (shorten the first note of an ascending leap and lengthen the first note of a descending leap), and Final-Ritard (ritardando at the end of the piece). Finally, the tonal tension is modeled applying High-Loud, Melodic Charge and Harmonic Charge (with the same quantity q = 3).

The dashed-dotted (red) lines in Fig. 3 refer to the new formulation of DM. Here, each accent marked in Fig. 1 is modeled by means of two new functions, for timing (duration) and dynamics (sound level) variations in the vicinity ot the accent. Each function admits five free parameters: the event peak, the

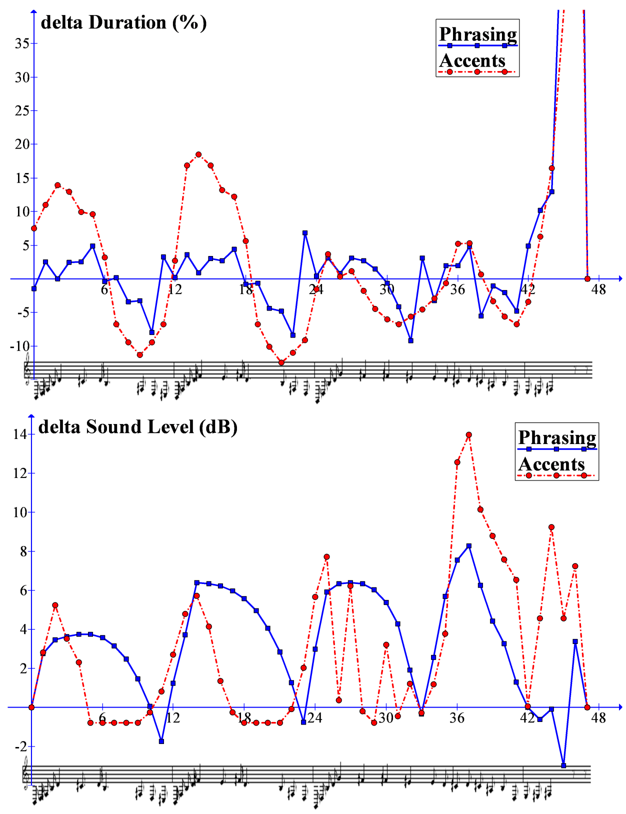

Figure 3: Example of mathematical modeling of timing (upper panel) and dynamics (lower panel) corresponding to the analysis of Chopin Prelude op. 28 n. 6 shown in Fig. 1. Solid (blue) lines are obtained applying the system of rules in the current formulation of Director Musices, that is dominated by phrasing (details in the text), while dashed-dotted (red) lines refer to the new formulation of DM, based on the model of accents.

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 saliences indicated by the analysis of Fig. 1. As in the previous formulation of the code, 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. As in the previous case, we chose the best settings of new rules, according to a preliminary auditory evaluation based on our ears. No rule belonging to the previous formulation is employed, apart from Leap-Tone-Duration and Final-Ritard (both independent on the context). Therefore a comparison between the two formulations is possible, as discussed in the next session.

4. CONCLUSIONS

Fig. 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. The opposite does not occur: DM' phrasing algorithms cannot reproduce all timing and dynamic fluctuations at the local note level. 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). DM has a mixture of top-down and bottom-up components (local and global); in its respect our approach is similar, but our simulations suggest that bottom-up approach can work alone, which would make the model more parsimonious. In this way, different subphrases can be modeled independently from one another, leading to higher variability in the profiles of timing and dynamics and hence a wider spectrum of performances.

5. ACKNOWLEDGMENTS

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

6. REFERENCES

[1] Clynes, M., “Sentics: The touch of emotions”, Doubleday Anchor, New York, 1977.

[2] Friberg, A., and Sundberg, J., “Does music performance allude to locomotion? A model of final ritardandi derived from measurements of stopping runners”, Journal of the Acoustical Society of America, Vol. 105, Issue 3, 1999, pp. 1469-1483.

[3] Friberg, A., Bresin, R., and Sundberg, J., “Overview of the KTH rule system for musical performance”, Advances in Cognitive Psychology, Vol. 2, Issue 3, 2006, pp. 145-161.

[4] Lerdahl, F., and Jackendoff, R, “A generative theory of tonal music”, MIT Press, Cambridge, MA, 1983.

[5] Parncutt, R., “Harmony: A psychoacoustical approach”, Springer-Verlag, Berlin, 1989.

[6] Parncutt, R., “Accents and expression in piano performance”, Perspektiven und Methoden einer Systemischen Musikwissenshaft (Festschrift Fricke), Peter Lang, Frankfurt/Main, Germany, pp. 163-185, 2003.

[7] Repp, B. H., “Diversity and commonality in music performance: An analysis of timing microstructure in Schumann’s Träumerei”, Journal of the Acoustical Society of America, Vol. 92, 1992, pp. 2546-2568.

[8] Sloboda, J. A., “Music structure and emotional response: Some empirical findings”, Psychology of Music, Vol. 19, 1991, pp. 110-120.

[9] Sundberg, J., “Computer synthesis of music performance”, Generative Processes in Music, Clarendon, Oxford, 1988.

[10] Sundberg, J., Friberg, A., and Bresin, R., “Attempts to reproduce a pianist’s expressive timing with Director Musices performance rules”, Journal of New Music Research, Vol. 32, Issue 3, 2003, pp. 317-325.

[11] Todd, N. P., “The dynamics of dynamics: A model of musical expression”, Journal of the Acoustical Society of America, Vol. 91, 1992, pp. 3540-3550.

[12] Todd, N. P., “The kinematics of musical expression”, Journal of the Acoustical Society of America, Vol. 97, 1995, pp. 1940-1949.