

A PRELIMINARY COMPUTATIONAL MODEL OF IMMANENT ACCENT SALIENCE IN TONAL MUSIC

Richard Parncutt

Centre for Systematic Musicology
University of Graz, Austria
parncutt@uni-graz.at

Erica Bisesi

Centre for Systematic Musicology
University of Graz, Austria
erica.bisesi@uni-graz.at

Anders Friberg

CSC, Dept. Speech, Music and Hearing
Royal Institute of Technology, Stockholm
afriberg@kth.se

ABSTRACT

We describe the first stage of a two-stage semi-algorithmic approach to music performance rendering. In the first stage, we estimate the perceptual salience of immanent accents (phrasing, metrical, melodic, harmonic) in the musical score. In the second, we manipulate timing, dynamics and other performance parameters in the vicinity of immanent accents (e. g., getting slower and/or louder near an accent). Phrasing and metrical accents emerge from the hierarchical structure of phrasing and meter; their salience depends on the hierarchical levels that they demarcate, and their salience. Melodic accents follow melodic leaps; they are strongest at contour peaks and (to a lesser extent) valleys; and their salience depends on the leap interval and the distance of the target tone from the local mean pitch. Harmonic accents depend on local dissonance (roughness, non-harmonicity, non-diatonicity) and chord/key changes. The algorithm is under development and is being tested by comparing its predictions with music analyses, recorded performances and listener evaluations.

1. INTRODUCTION

The music rendering competition Rencon (Hiraga et al., 2004) has shown how difficult it is to simulate the expressive performance of familiar Western classical music. Despite decades of research, automatically generated musical expression is often unconvincing - even in shorter musical excerpts, in which relatively intractable contextual factors such as genre-specific expressive devices and the music's programmatic meaning can be neglected in a first approximation.

Our approach is based on the analysis-by-synthesis approach of Sundberg (1988) and Friberg (1991), and their rule-based performance rendering system *Director Musices*. Like them, we begin with the score and adjust the timing and dynamics of an automatically generated performance by applying rules to selected structural features.

Inspired by Sundberg and Friberg, Parncutt (2003) developed a new theoretical foundation. In a broad definition, an *accent* is any musical event that seems important or attracts the attention of a listener. An *immanent* accent is an accent that is determined by the musical structure as suggested by the musical score; a *performed* accent is added to the music by the performer by manipulating dynamics, timing, articulation or timbre. In both cases, the perceptual *salience* of an accent is its perceptual or

subjective importance, or the degree to which it attracts a listener's attention.

Many of the performance rules of Sundberg and Friberg can be reinterpreted in terms of this general concept of accent. The accent concept unifies the theory under a new general umbrella and establishes a stronger connection between performance rendering and the academic discipline of music theory and analysis.

The accent approach is ultimately based on the psychoacoustics of musical event perception. The noteheads in a musical score may be equal in size, but when the music is performed, they do not sound equally important. Imagine that a score is performed completely deadpan, with timing corresponding exactly to the score and all tones played at a sound level that would make them equally loud if played alone. In context, those tones will seem to differ in loudness, because they mask each other to different extents (Egan & Hake, 1950). On average, the outer voices typically seem louder or clearer than the inner voices since the (fundamental frequencies of the) outer voices are primarily masked on one side, but the inner voices are masked on both sides.

Now imagine that the performer adjusts the loudness of the tones so that they are equally loud in spite of masking. The tones will still seem unequally important due to the musical structure (phrasing, meter, melody, harmony). Because these effects arise initially from the structure and not from the performance, they may be considered immanent to the score.

Immanent accents may be divided into several kinds, which Parncutt (2003) identified as *grouping*, *metrical*, *melodic* and *harmonic*. Note that Lerdahl and Jackendoff's (1983) term *structural accent* has essentially the same meaning as Drake and Palmer's (1993) *grouping accent*. We adopt the latter; Lerdahl and Jackendoff's term is inappropriate insofar as all immanent accents may be considered "structural". The term "grouping accent" is also problematic, since a group of tones or sound events can be either *serial* (like a phrase) and *periodic* (like a metrical pattern or pulse); but the alternative term "phrasing accent" is equally problematic, because it may be misleading to describe a whole piece as one long "phrase" (the word "phrase" suggests a time period equivalent to one breath, while singing or playing an instrument).

Everyday usage of the word "accent" suggests that all notes in a score can be classified either as accented or unaccented - just as all tones, intervals or chords can be classified as consonant or dissonant. In fact, both conso-

nance and accentuation vary on a continuous scale. The salience of an immanent accent may be compared with its perceptual importance when the music is heard in a dead-pan performance - before the performer manipulates accent salience with performed accents.

On this basis, Parncutt (2003) developed a new approach to modeling expression (timing, dynamics, articulation, timbre and so on) in Western classical music. The procedure involved first estimating the perceptual salience of the main immanent accents in the score, and second exaggerating or bringing out those accents in a computer-generated performance. Performers tend to clarify or disambiguate the structure of a piece of music for the listener by adding extra salience to selected immanent accents (Bisesi et al., 2012). Common expressive manipulations include slowing the tempo (or adding extra time) and increasing the loudness in the vicinity of immanent accents. It is also possible to unexpectedly reduce loudness to attract attention to an event. Important events are typically delayed, played louder, or both. Of course there are exceptions, but they happen less often.

In order to model such effects convincingly, we must address several questions. According to what general principles can immanent accents be identified? What general principles determine their salience? How are tempo, dynamics and other parameters typically varied in the vicinity of accents? Over what time period before and after an accent are they varied? What is the shape of the timing or dynamic curve leading up to and away from the accent? This proceedings contribution addresses the first two questions by identifying the main principles and sketching a computer algorithm to identify and evaluate immanent accents.

2. GROUPING ACCENTS

A phrase is a temporally contiguous series of tones or musical events that are grouped in our perception by the Gestalt principle of temporal proximity. Grouping accents occur at the start and end of phrases or sections. The listener's attention is drawn to structural boundaries for the simple reason that they delineate the structure; we assume that a listener only "understands the music" if s/he intuitively parses the structure "correctly", i.e. more or less as a composer, performer or theorist would do. The first note of a phrase or section is important because it announces something new; the last note is important because it announces the end of a group (closure). Performers tend to slow at phase boundaries; listeners consequently expect these tempo fluctuations (Repp, 1998). Bisesi et al. (2012) found a general agreement about the position of phrase beginnings, endings and climaxes across performances and listeners.

Grouping accents often coincide with other accents. Consider for example metrical accents. If the start of a phrase coincides with the start of a measure, metrical and grouping accents coincide. We may assume that a compound accent is created, whose salience is greater than that of each of the two accents considered separately. If phrasing and metrical accents do not coincide, there is an upbeat or anacrusis. In that case, we may hear the start of the phrase as more or less important than the following

upbeat – depending e.g. on how we are listening, or whether we are singing or playing percussion.

To analyse grouping accents in a score, it is necessary first to parse the piece according to its hierarchical phrasing structure. The outcome is sometimes ambiguous: different theorists might offer different analyses for the same passage. For this reason, we have not attempted an automatic analysis here. Instead, we describe the basic subjective principles according to which a theorist might segment a passage of music into phases and group these together to make a hierarchical structure. A semi-algorithmic approach of this kind was also adopted by Lerdahl and Jackendoff (1983).

The first step is to regard the entire piece or excerpt as one long phrase. The next is to divide this long phrase into a number (normally 2 or 3) of subphrases of nominally equal importance. Then divide each subphrase into sub-subphrases, and so on until arriving at the level of individual notes. This process is similar to the process of parsing speech utterances or written text in linguistics.

How exactly is the subdivision decided at each level? Musicians and theorists know intuitively how to do this; if we are to create a more objective procedure, we need to make those intuitive processes explicit. First, consider inter-onset intervals (IOIs) – the time interval between the onset of a note and the onset of the next note. The IOI of the last note in a phrase (i.e. the IOI between that note's onset and the onset of the first note in the next phrase) is often relatively long by comparison to IOIs within the phrase. Moreover, rests are more likely to occur between phrases than within them. Second, there might be a relatively large leap in pitch between the last note of a phrase and the start of a new one. Third, phrases often rise in pitch near the start and fall in pitch near the end, so a parsing that produces such phrase shapes may be preferred. Fourth, if a melody includes repetitions of recognizable motives, the phrasing should not break these motives up.

A final rule to observe is that subphrases of a given phrase should have roughly equal length; the longest subphrase should in any case not be more than twice as long as the shortest subphrase. If this rule is broken, the problem can be solved by joining together two shorter subphrases and dividing that phrase again at the next level down the hierarchy.

The phrasing structure of a piece may be ambiguous – different interpretations are possible and may even seem equally valid. In this case, different possible interpretations may be considered separately, and the validity of each interpretation may be estimated quantitatively in a comprehensive algorithm. We might e.g. consider interpretation A to be 70% valid and interpretation B to be 30% valid, and retain both alternatives in further analyses or applications.

Once the phrasing structure has been determined, the grouping accents can be located and their salience estimated. A grouping accent occurs at the start and end of a phrase at any hierarchical level. In a first approximation, the salience is simply the number of levels at which a given event marks the start or end of a phrase (cf. Todd, 1985). In a more sophisticated approach, the salience of a

grouping accent may be the sum or another combination of such salience values.

In musical scores, phrasing is often explicitly marked by the composer, arranger or editor. Examples of phrasing and corresponding accents are shown in Figures 1a and 1b. In both cases, we have simply followed Chopin's phrase marks. In Figure 1a, the passage has been divided into two phrases. The relatively large time gap between the phrases and the similarity of the two phrases makes the phrasing quite unambiguous. Because the two phrases combine into one phrase at the next hierarchical level, the grouping accent at the start of the first phrase is stronger than that at the start of the second phrase. The start of the first phrase is also the start of phrases at two higher levels (groups of four and eight phrases respectively), which accounts for the large difference in salience between the two grouping accents. The grouping accents at the ends of the phrases are determined by the same logic in reverse: the biggest phrase-ending grouping accent occurs at the end of the piece. Figure 1b spans a single phrase marked in the score, so it begins and ends with a grouping accent. The phrase is difficult to divide unambiguously into two or three shorter phrases, so we treat Chopin's phrasing as the lowest (fastest) level. In an analysis of the whole piece, we would have marked further phrases at higher (slower) hierarchical levels.

3. METRICAL ACCENTS

Like the phrasing, the meter of a piece of music usually has a hierarchical structure. Consider a piece in $\frac{3}{4}$ time. There is a basic $\frac{1}{4}$ note pulse; $\frac{1}{4}$ notes are grouped into threes making a $\frac{3}{4}$ note pulse at the barline. These are two adjacent levels of a hierarchy. The hierarchy can be extended both upwards and downwards: if note values smaller than a $\frac{1}{4}$ note are used, they can create faster pulses, and if groups of measures produce new, perceptible periodicities, they can be regarded as slower pulses called *hypermeter*.

In Figure 1a, there is a bigger metrical accent at the start of the first measure than at the start of the third measure, and both these accents are stronger than at the start of measures 2 and 4. We already encountered this 4-1-2-1 pattern in hierarchical phrasing analysis. However is not always clear whether these groupings are being perceived as phrases or hypermeter. For the purpose of performance rendering, phrases are already accounted for. For these reasons we do not consider higher-level hypermeters. But we do consider groups of two measures. For example, the metrical accent at measure 1 of Figure 1b is greater than the metrical accent at measure 2.

If a piece remains in the same time signature throughout and there are no obvious ambiguities (i.e. if the composer evidently intends the listener to perceive the notated meter), then the analysis of metrical accent within measures is straightforward. The salience of the metrical accents can be estimated in the same way as the grouping accents were determined above. In a first approximation, the salience of a metrical accent is the number of different level of pulsation to which it belongs. In a second approximation we include the dependence of pulse salience on tempo (pulses near about 100 per minute are the most

salient) and add the salience of the pulses to which each event belongs.

If the piece stays in the same meter and has little or no syncopation, we can establish a complete metrical hierarchy. Any change in the meter causes a temporary weakening of the hierarchy as the new meter is established. Syncopations typically make other interpretations possible. This has not yet been modeled, but a systematic approach might be to present the different possible structures and weight them relative to each other: interpretation A might be the more likely with a probability of 70% and interpretation B with 30%.

Our algorithm currently works as follows. We first mark four metrical levels. The note value assigned to each beat level is given in Table 1 for the most common time signatures. The table could easily be extended to other time signatures. The conventional "beat" generally corresponds to level 1 in the table. The barline corresponds to level 2 for simple metres in which the measure is 2 or 3 beats, and to level 3 for compound metres in which the measure is 4 or 6 beats.

Time signature	Metrical level			
	Level 0	Level 1 (beat)	Level 2	Level 3
4/4	1/8	1/4	2/4	4/4
2/2	1/4	1/2	2/2	4/2
4/2	1/4	1/2	2/2	4/2
2/4	1/8	1/4	2/4	4/4
$\frac{3}{4}$	1/8	1/4	$\frac{3}{4}$	6/4
3/8	1/16	1/8	3/8	6/8
6/8	1/8	3/8	6/8	12/8
9/8	1/8	3/8	9/8	18/8

Table 1. The period of each metrical level expressed as note values for different time signatures.

Next, we compute the salience of each metrical level. Following Parncutt (1994), we assume that the function of pulse salience against period is a Gaussian function relative to a logarithmic scale of period:

$$\text{Salience}_i = e^{-0.5 \left(\frac{\log X - \log M}{\log S} \right)^2},$$

where X is the period of the metrical level in seconds, M = 1.0 seconds is the centre (mean) of the Gaussian distribution, S = 1.65 is the standard deviation of the distribution, and i is the metrical level (0..3). In Parncutt (1994) the mean M was smaller (0.6..0.7 seconds); we found that increasing it improves modeling of hypermetre.

Finally, we calculate the metrical accent salience of each point in time in the score. It is simply the sum of the salience of all metrical levels including that note.

4. MELODIC ACCENTS

Melodic accents are marked "C" in Figure 1. The C stands for contour (or melodic contour accents) and avoids confusion with M for metrical accents. For an overview of research on melodic accent, see Huron and Royal (1996).

In Figure 1a, the first melodic accent is at the start of measure 1. The accent is evidently due to the (rising) leap before the accent, which attracts attention to it. The next two marked accents also follow rising leaps. Figure 1b also shows examples of melodic accents in the bass line. In all these cases it appears that the salience of the accent is due to a combination of two factors: the size of the leap preceding the accent, and the distance of the accent from the centre of the melody's range or ambitus. Further principles determining melodic accents appear to be the following: only local peaks or valleys, and only tones following leaps (3 semitones or more), can bear a melodic accent; melodic accent salience depends on a combination of leap size and distance from local mean pitch; the first and/or last in a group of repeated notes may be accented; and melodic peaks generally receive stronger accents than melodic valleys.

Our computer implementation works as follows. First, the mean pitch is calculated for each track individually. Then each tone is assigned a salience S_1 for the pitch deviation from the mean:

For notes above the mean:

$$S_1 = |\text{interval from mean in semitones}|$$

For notes below the mean:

$$S_1 = |\text{interval from mean in semitones}| * 0.7$$

Then each tone is assigned a salience S_2 according to the size of the preceding interval:

For rising intervals:

$$S_2 = |\text{preceding interval in semitones}|$$

For falling intervals:

$$S_2 = |\text{preceding interval in semitones}| * 0.7$$

The final value for melodic salience = $(S_1 + S_2) / 15$.

5. HARMONIC ACCENTS

Harmonic accents are marked "H" in the figures. To begin again with some examples: The harmonic accent in measure 1 of Figure 1a is due to the (mildly) dissonant 6th interval above the root, which resolves to the consonant 5th. Since this is a rather weak dissonance, the estimated salience of the harmonic accent is rather low. The accent in measure 3 is due to the diminished triad, two of whose tones do not belong to the prevailing diatonic scale. The accent depends partly on the dissonance of the diminished triad (independent of context) and partly on the double departure from diatonicity. In Figure 1b, the first harmonic accent announces the start of a new harmonic region. The preceding passage is in F# major; the D# minor chord heralds a passing transition to the key of C# major followed by a sequential repetition that suggests B major. Later harmonic accents are due to the relative dissonance of specific chordal sonorities.

These examples suggest that harmonic accent has several components. First, the dissonance of a chord (considered in isolation, but also relative to the dissonance of preceding and following chords) may attract attention and hence produce an accent. This is difficult to formulate in an algorithm since there is no accepted general model for the dissonance of a sonority in western music (Parncutt & Hair, 2011). If we assume that the

dissonance of a sonority is a combination of its roughness and (lack of) harmonicity, dissonance could be estimated most simply by counting the number of clearly dissonant intervals (minor seconds, tritones, major seconds) and harmonicity could be estimated by the presence of clearly harmonic intervals such as perfect fifths and fourths, or the salience of the root according to Parncutt (1988). But there are further complications: a chord may be merely implied, and implied chords are often ambiguous. At the start of measure 1 of Figure 1a, do we have an inverted A-major chord or a suspension above an E-major chord? A simple algorithm is more likely to predict the former, whereas a (Schenkerian) theorist will indicate the latter.

Second, harmonic accents in major-minor tonal music are produced by tones foreign to the prevailing key. If the key of a passage is relatively clear, the salience of this kind of accent can be predicted using the key profiles of Krumhansl and Kessler (1982). These profiles may be considered as quantitative estimates of the harmonic stability of each tone in the chromatic scale in a given major or minor key. The lower the stability of a tone, the greater the harmonic accent at that tone. The harmonic accent of a chord may be estimated by combining accents for individual tones.

Third, harmonic accents are produced by important harmonic shifts. This aspect could be modeled by a key-tracking algorithm. Where modulations are predicted to occur, the chord announcing the modulation may be accented. But this procedure may not work for pivot chords, which belong to both a preceding and a following key; and music theorists differ markedly in their interpretation of modulations. At one extreme, any accidental may be considered to suggest a modulation, while at the other extreme, a whole extended piece may be considered to stay in the same key in spite of extensive chromaticism (chromaticisms may be instead function as tonicizations). This theoretical debate can be avoided by focusing on performance expression in modulating passages: if a performer brings out a modulation, it exists. The theoretical debate about modulation versus tonicization could be resolved by considering "real music" rather than the score.

We have not yet implemented the above approach. For the moment we are using the existing approach of Director Musices. The current implementation requires that a functional harmonic analysis is manually provided in the score. The salience of a harmonic accent is computed at each chord change as follows:

$$\text{Salience} = 1.5 * \sqrt{\text{Harmonic charge}}$$

Harmonic charge is a measure of the tonal perceptual distance of the chord from the prevailing key; see Friberg (1991) for a technical description.

6. EXAMPLES

In Figure 2 and we have tentatively applied the algorithm to the passages illustrated in Figure 1.

7. CONCLUSION

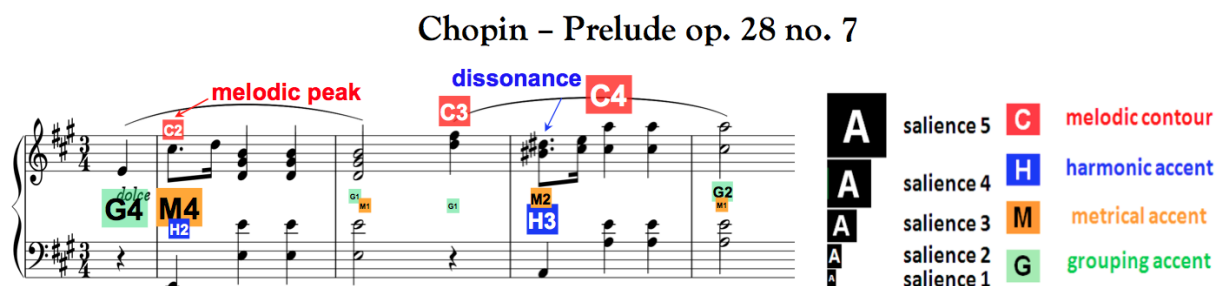
This has been a preliminary sketch of the main principles behind a new algorithm to estimate the perceptual salience of immanent accents in tonal western music - a step toward a new semi-algorithmic approach to performance rendering. We say “semi-algorithmic” because we are currently cautious about completely automatizing the procedure. Even when our algorithms for predicting the salience of different kinds of accents are refined and consistently make feasible predictions, some details of the algorithms will remain dependent on style. We also anticipate that the relative importance of different kinds of accents will depend on stylistic context.

We are testing versions of the algorithm in several ways. First, we are comparing its predictions with our music-theoretical intuitions, and judging the musical naturalness of the resulting performance renditions (analysis by synthesis). Given the large number of options at the beginning of such a project and the impossibility of considering all options systematically, this is the most practical way to proceed. We are then comparing predictions of a prototype with analyses of music theorists and our analyses of expression in recorded performances, and making improvements based on the data (cf. Thompson et al., 1989). Finally, the “musicalness” (musical quality, expressive content) of performances generated by the algorithm will be tested in listening experiments in which expert listeners judge the quality of performance renditions.

8. REFERENCES

- Bisesi, E., MacRitchie, J. & Parncutt, R. (2012). Recorded interpretations of Chopin Preludes: Performer’s choice of score events for emphasis and emotional communication. In E. Cambouropoulos et al. (Eds.), *Proceedings of the 12th International Conference on Music Perception and Cognition* (pp. 106-107).
- Drake, C., & Palmer, C. (1993). Accent structures in music performance. *Music Perception*, 10, 343–378.
- Egan, J. P., & Hake, H. W. (1950). On the masking pattern of a simple auditory stimulus. *Journal of the Acoustical Society of America*, 22, 622-630.
- Friberg, A. (1991). Generative rules for music performance. *Computer Music Journal*, 15 (2), 56–71.
- Hiraga, R., Bresin, R., Hirata, K., & Katayose, H. (2004). Rencon 2004: Turing test for musical expression. In *Proceedings of the 2004 conference on New interfaces for musical expression* (pp. 120-123).
- Huron, D., & Royal, M. (1996). What is melodic accent? Converging evidence from musical practice. *Music Perception*, 13, 489-516.
- Krumhansl, C. L., & Kessler, E. J. (1982). Tracing the dynamic changes in perceived tonal organization in a spatial representation of musical keys. *Psychological Review*, 89, 334-368.
- Lerdahl, F., & Jackendoff, R. (1983). *A generative theory of tonal music*. Cambridge, MA: MIT Press.
- Parncutt, R. (1988). Revision of Terhardt’s psychoacoustical model of the root(s) of a musical chord. *Music Perception*, 6, 65-94.
- Parncutt, R. (1994). A perceptual model of pulse salience and metrical accent in musical rhythms. *Music Perception*, 11, 409-464.
- Parncutt, R. (2003). Accents and expression in piano performance. In K. W. Niemöller (Ed.), *Perspektiven und Methoden einer Systemischen Musikwissenschaft* (pp. 163-185). Frankfurt/Main, Germany: Peter Lang.
- Parncutt, R., & Hair, G. (2011). Consonance and dissonance in theory and psychology: Disentangling dissonant dichotomies. *Journal of Interdisciplinary Music Studies*, 5 (2), 119-166.
- Repp, B. H. (1998). Variations on a theme by Chopin: Relations between perception and production of timing in music. *Journal of Experimental Psychology Human Perception and Performance*, 24, 791-811.
- Sundberg, J. (1988). Computer synthesis of music performance. In J. A. Sloboda (Ed.), *Generative processes in music*. Oxford: Clarendon.
- Thompson, W. F., Sundberg, J., Friberg, A., & Frydén, L. (1989). The use of rules for expression in the performance of melodies. *Psychology of Music*, 17, 63-82.
- Todd, N. P. McA. (1985). A model of expressive timing in tonal music. *Music Perception*, 3, 33-58.

Figure 1. Subjective analysis of immanent accents and their salience in music of Frédéric Chopin. (a) The first four measures of Prelude Op. 28 No. 7. (b) The first two measures of the central section of Prelude Op. 28 No. 13.



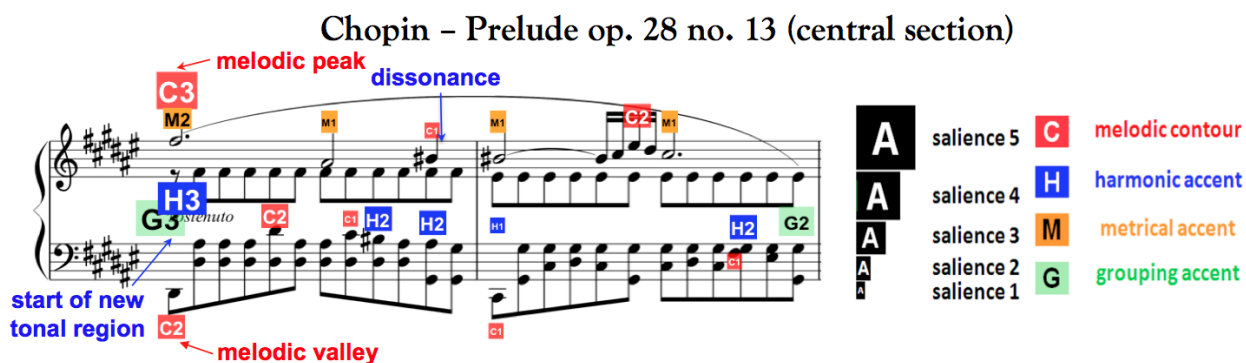


Figure 2. Model predictions for Prelude Op. 28 No. 7 and No. 13 (central section) by Frédéric Chopin.

