MADELEINE LE BOUTEILLER
01/03/2015 - 31/07/2015

---

# Expressivity in Debussy's Sonata for Cello and piano
## Towards an annotation system of cello expressiveness

---

**Summary**: Each interpretation of the same piece of music is unique: every performer has his own expressivity. Some features of expressiveness in played music, such as tempo variations and loudness, can be measured and extracted from recordings. Working on chosen excerpts of Debussy's Sonata for cello and piano, expressive parameters have been extracted from several recordings, and an annotation system has been developed. Augmented scores have been generated, where we can read the expressive features below the notes. Combined with cognitive models, that predict our understanding of melodies and the feeling of tension in tonal music, we could find out some correlations between these models and the interpretation parameters: cellists react to the dissonance of the music they play, to the way melodies continue in an attended or unexpected way. The implementation of these models has been done, so as to get an automatic calculation of the models.

**Keywords**: *Music, interpretation, expressiveness, annotation, computation, musicology*

Internship supervised by:
**Elaine Chew**
elaine.chew@qmul.ac.uk

Queen Mary University of London
Mile End Road
London E1 4NS

August 7, 2015

# Contents

# Introduction

Music from the beginning of the twentieth century often tends to escape the traditional forms of analysis. This is the case of Debussy's sonata for cello and piano, composed in 1915. The complexity of the piece is a stumbling block to analysis. In some moments tonal, atonal or more modal, it is neither globally tonal, atonal or modal, and it completely escapes functional analysis [1]. Based on this observation, the idea of the internship is to get round of the traditional analysis. We will attempt to bring some material of understanding in applying some cognitive models of music perception, relied with an analysis not of the written music, the score, but drawn from performances, from recordings.

As analysis can not be provided successfully by the classical analysis theories, we attempt here to provide an alternative to classical music analysis methods. Score and performance are two representation of music: the score is the representation of the music the composer thought, and the recording is a good approximation of what is given to the listeners. Each interpretation of one piece is unique, several performers can play the same piece in many different ways. Even if they read the same score, musicians can convey many different interpretations. The way they segment the piece, the notes or phrases they choose to stress by varying tempo, dynamics and timbre, are features that make one interpretation different from another one. An other atmosphere in conveyed by every new interpretation of the work.

The project reported in this document is to extract some expressive features from the performances and mark them up into the score. We will go from the recordings of the sonata, so as to pick up some elements of the expressiveness conveyed by the performers, and go back to the score to automatically annotate it with those parameters. We are then providing some augmented scores which contain a representation of the interpretation of one performer.

The way a musician decides to interpret a work depends on his own understanding of the piece. It is a sometimes unconscious, sometimes conscious, sometimes learnt process, which is actually never definitive. Music theoreticians, cognitive scientists, have developed some models of our understanding of tonal music. The tonal tension model models the variations in stability and relaxation in tonal music [2]. It quantifies the tension arisen from the harmony, melody and structure in music. The Implication-Realisation model, developed by Narmour, quantifies our expectations when we hear a melody [3]. Combining ideas from these two models, we can correlate the expressivity drawn from our recording analysis with an implicit understanding of music.

So as to do that, I focused on one eight bars theme from the first movement of the sonata, which is repeated with variations in the recapitulation, by the end of the first movement (Fig. 12). This theme is divided into two four bar phrases, which first two bars are the same for the cello. It allows comparisons across themes and performers.

I have used in this study five recordings of Debussy's sonata, those of cellists Paul Tortelier, Maurice Gendron, Gregor Piatigorky, Mischa Maisky, and Susanne Beer, with their respective pianists C.

This project can be the beginning of a language. Like it already exists a system for annotating prosody in speech with the system ToBI [4] [5], nothing has been done yet or the annotation of interpretation in music.

I will first present the state of the art in the fields of speech annotation, expressive performances analysis, and cognitive models of musical understanding (1). Then, I will present the annotation system I developed (2), as well as its implementation (3), before finally showing and discussing some interesting results (4).

# 1 Literature review

## 1.1 Expressive analysis and cello performances

Todd built a model of musical expression, based on the assumption that a phrase is often shaped by a crescendo on the beginning and decrescendo on the end, with the corresponding timing variation from slow to faster to slower on the end [6]. He built an quantitative model of the relation between tempo and dynamics.

Indeed, phrases boundaries can often be automatically identified, as slowing down by the end of a phrase is a largely shared parameter of expressivity [7].

Timing and dynamics in cello performances have been investigated by Ju-Lee Hong [8]. She studied the relation between tempo and loudness variation in Bach's *Sarabande* in C major, from the third Suite for cello solo. Her aim was to find out whether cellists play louder when faster, and lower when they slow down. She showed that this was not how performers use to play, as Todd assumed, and that their expressivity depends on their own understanding of the *Sarabande*, and how they want to shape it.

Parameter of expressivity have already been studied in several ways. If dynamics and timing are the most recognized and obvious features of expressivity, we have to add, for string instruments, the use of vibrato.

## 1.2 Annotation models

### 1.2.1 ToBI: a speech annotation system

ToBI (Tones and Break Indices) is a transcription system for speech [5]. As a text can convey several different meanings, depending on the way it is said, prosody in speech is crucial for the understanding of the intended meaning of an utterance. ToBI is an an agreed system for the annotation of prosody, relying the acoustic speech signal with the text and the discourse structure. The aim is to develop and share a prosodically transcribed data base, using the same annotation conventions, as the understanding of prosody is crucial for research in natural language processing, speech synthesis, and the development of spoken language understanding systems.

ToBI annotates the intonation patterns, breaks in the oral flow, an and other aspects of the

prosody, with a well-defined dictionary of annotation symbols [4]. ToBI annotates the different levels of breaks (between words, sentences), the phrase accents (high or low in pitch, with pitch variations). It adds to this an orthographic layer, and a miscellaneous tier (for annotating laugh for instance). The annotation process consists in listening to the utterances and marking them manually.

Relation between written music and played performance is similar to the one that binds written text and oral discourse. The score and the performance are two different representations of music, as written and spoken discourse. There is no annotation system that does something equivalent of ToBI for music.

### 1.2.2   Expressive markings for music generation

Chrisopher Raphael developed an annotation system for expressive melody synthesis [9]. Focusing on folk-like melodies, a small set of symbols indicates for each note its role in the context, in therms of stress and direction. For instance, we can distinguish notes that lead forward or recede the movement, notes that are stressed as point of arrival. From an annotated score, the aim is to synthesise an electronic melody with expressivity. By varying pitch (notes changing, vibrato and glissendo) and intensity (dynamics), an expressive melody can be generated. Two functions, frequency and intensity, are built from the annotated score, so as to produce the audio signal.

In this case, the labelling is done manually and it reflects the expressivity the annotator wants the melody to convey.

## 1.3   Cognitive models of musical understanding

### 1.3.1   Tonal Tension Model

The experience of tonal music listening consists for a part in the feeling of raises in tension and relaxation. In the realm of a tonality, dissonant chords and key changes convey a feeling of destabilization, whereas a tonic chord brings relaxation. This refers to the tonal tension's variations. It arouses an emotional response. F. Lerdahl and C. Krumhansl have developed a model to quantify it [2] [10], based on ideas of the Generative Theory of Tonal Music [11]. The idea of *tension* is being implicitly understood as an unstable state which tends to relax into a more stable one. If we can see many musical components such as rhythm, dynamic, orchestration and texture, that can contribute to the global musical tension, the tonal tension is one of them (in the realm of tonal music only), made of a notion of dissonance (both sensitive and cognitive dissonance: from a psychoacoustic and a music theoretical point of view). It has been shown to shape some aspects of the listening experience, as well as of the performances [12].

The tension is an additive combination of three factors:

- the distance between chords within a tonal region: two chords are considered close to each other of distant, according to several factors: the distance between them in the circle

of fifths, and the number of non-common tones between them. The model assumes that a move from one chord to a very different one is a raise in tension. Distances can be calculated in a sequential way (chord after chord) or in a hierarchical way (more likely to account for large scale structure): events are put in relation according to their structural importance. This way requires a reduction, close to a Schenkerian analysis.

- the **surface tension**, is a notion of local degree of dissonance. It quantifies how dissonant an event is (a chord, or a chord with a passing tone in the melody), considering the inversion of the chord, the ton in the upper voice, and the non-harmonic tones present in the chord (Fig. 5). A detailed calculation is presented in 3.1.2.

- the **melodic attraction**: within a tonal region, some tones tend towards others: in C Major, B is strongly attracted by C. Depending on the stability of notes within a key and the distance (in semi-tones) between two notes, is built the melodic (or voice-leading) attraction (Fig. 13). Its calculation will be explained in 3.1.1.

This tonal tension model considers only the harmony. Rhythm, timbre, dynamic, also contribute to the global musical tension. This model tries, with theoretical concepts, to fit the empirical feeling of listeners of tension-relaxation. It has been tested over listeners [10] [13] and shown to be able to account for some interpretation parameter, especially the tempo [12].

Performers may emphasize passages or events with a high degree of tension.

### 1.3.2   Narmour and the Implication-Realization model

Narmour's model focuses on melodies, independently of style or tonality. Its aim is to model our expectation in hearing a succession of tones. The ground idea of this model is that the denial of an expectation causes emotional effects by the listener. Emotion in music is being aroused "when an expectation activated by the music stimulus is temporary or permanently blocked" (Narmour). His theory is based on some principles, arisen from the *Gestalt* theory: it explains how we naturally tend to group together similar elements or stimuli (both auditory and visually). They are innate and abilities. The Implication-Realization model is a theory of our understanding of melody, based on perception and cognition [3] [14]. It describes our cognitive response to melodic events: as we hear an interval, we build some expectation for the coming tone. The model quantizes this expectation.

The theory considers three tones. Narmour came to formulate five principles to predict how the third tone continues well the melody initiated by the first two tones. For instance, if we hear a large interval going up (let's imagine C - A), we will be more likely to expect the melody to go down with a small interval (G). So, C - A - G is considered more likely than C - A - E or C - A - C. Depending on the size and direction of the consecutive intervals, the five principles predict how well a tone will fit the expectancy of the listener.

The model describes tone to tone expectancies for continuation of melodies. It has been experimented over listeners [15] [16]. According to Narmour, violation of implication (of what

we expect) produces affective and aesthetic effects. Listeners as well as performers may react or play differently depending on how they expect one note.

## 1.4  Study of expressive interpretation

Caroline Palmer did study the relations between the two above presented models with timing and dynamics variations in piano performances [12]. Working on recordings of a Mozart's piano sonata, she was able to identify some correlations between the tension model and the expectation, and some expressive features: pianists were more likely to play slower passages with high degree of tension, and to play louder the unexpected notes.

# 2  Annotation system

The way a musician performs a piece is unique, and it is distinguishable from other performances by many fine parameters of playing. Through their rhythm, dynamics, motions, musicians emphasize some notes, phrases, they convey the movement, the direction they want to give to the music they play. They manage to show and express their own understanding of the score.

After the identification and extraction of some particular expressive features of the playing, we will incorporate them into the score. We obtain then an augmented score, containing the expressivity of one performance.

I analysed some chosen parameters of the interpretation, which are both interesting in terms of musicality, and reliable. To do that, I have built a small dictionary of appropriate symbols (Fig. 4). For this analysis, I focused on four four-bars long phrases (Fig. 12).

## 2.1  Timing parameters

Rhythm in performance is never completely regular and metronomic. Even if sometimes we can not explicitly identify the tempo variations, they are still audible and we can perceive the expressivity they convey.

Looking at a large and a middle scale, timing is a good indicator of phrasing. As performers usually slow down toward the end of a phrase [7]. At a smaller scale, it can be an indicator of stress: the musician may play a note longer if he wants to emphasize it.

I choose to indicate in the annotations the notes that are played longer than the mean tempo of the phrase. I assigned them a _ above these notes. They are a good indicator of where the tempo slows down.

I also annotated with a = the notes that are played at a slower tempo than their two neighbouring notes. This is a better indicator of local stress.

The detailed annotation process will be explained in 3.2.2.

## 2.2   Loudness parameters

The loudness is obviously an important parameter of the interpretation. Usually, some dynamics marking are already put in the score by the composer. It is especially true by Debussy, who used to write a lot of annotations about timing (lent, poco animando, cédez...) and dynamics (piano, forte, crescendo...). Even if those markings are numerous and precise, it remains a large freedom for the musician to play his own tempo and dynamics.

As loudness indicators, I decided to look at different details that are relevant in term of interpretation. First, the shape of the long notes is a freedom own by the cellist alone. The dotted crochets can be played in crescendo or decrescendo. We will put the information back in the score through the signs $<$ and $>$.

Another important parameter is the way cellists shape the three quavers before the end of the second bar of the phrase. Theses three-notes patterns can be played in very different ways, depending on which note the cellist wants to enhance: the one on the beat, or the first note of the pattern (Fig. 1) We will mark with a + the note of the three that is played louder[1].

I also decided to look at the different ways of starting the phrase. The attack note is a significant point of diverse ways to play and rendering expressiveness, as it can be firstly loud, and then decrescendo, or on the opposite, played in crescendo, with a smooth attack and the full sound coming later. It conveys either an impression of determination in the first case, or a feeling of smoothness in the second one. The score will be annotated with $>$ if the note is played in decrescendo, $<$ if in decrescendo. The mechanism will be detailed in 3.2.3.

## 2.3   Other parameters of expressiveness

Two other parameters are also relevant for the expressivity of a cello player. First, the vibrato. Short notes are usually played without vibrato, whereas long notes with vibrato (Fig. 2). In our excerpts, all the quavers are neither short enough to not have enough time to vibrate, neither very long. Cellists can choose how they play. I marked it with a ⌣.

Then, *portamento* is when we can hear a glissendo, a continuous pitch flowing from one note to the following one. They will be marked up with / or \, whether it goes up or down (Fig. 3).

# 3   Implementation method

I will present in this part the method I developed and used to apply the cognitive models and run the annotation process to Debussy's sonata. I used the library music21 to write programs in Python that can automatically extract features from the scores and the recordings, and annotate

---

[1]You may note that the piano plays an onset with the second quaver of the cello, for every pattern. The loudness is not the one of the cello alone, as its calculation has been done on the recordings of piano and cello together. The important loudness is indeed the global loudness conveyed by the duo and perceived by the listeners.

[2]Only the very last note of the phrase will not be considered in this study: it is always played in decrescendo by every cellist.
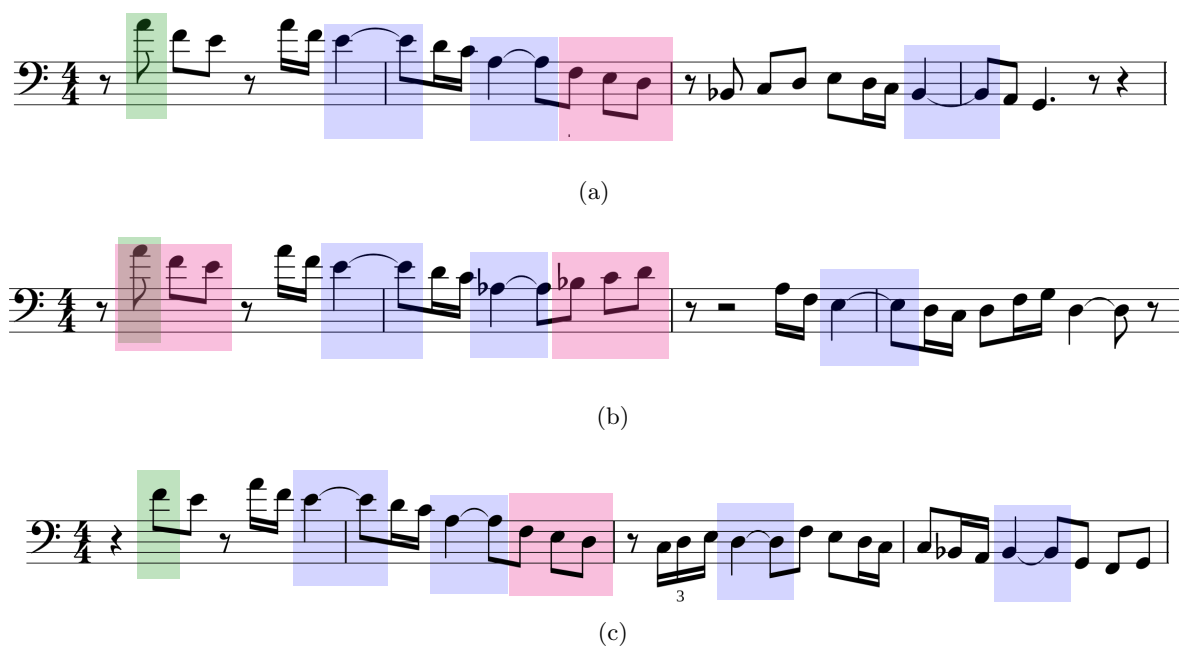
(a)



(b)



(c)

Figure 1: Cello part of the chosen excerpts. (a) and (b): first exposition of the theme. (a) and (c): second apparition of the theme, in the recapitulation. The first part of the theme is exactly the same for the cello, and the second half variates. In coloured rectangles, the notes which were subject to a loudness analysis: blue: the dotted crochets[2]; red: the three-quavers patterns; green: the attack note.
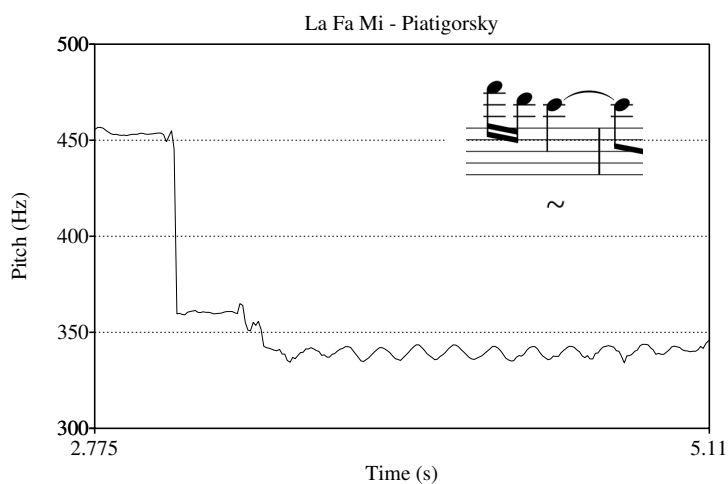


Figure 2: Fundamental frequency contour (la fa mi) extracted from one cellist's recording (Piatigorsky), with the software Praat, with the corresponding score excerpt. We can observe a large and regular vibrato on the E. Piatigorsky's cello is tuned very high: the A is above 450 Hz , instead of the usual 440 Hz.

7
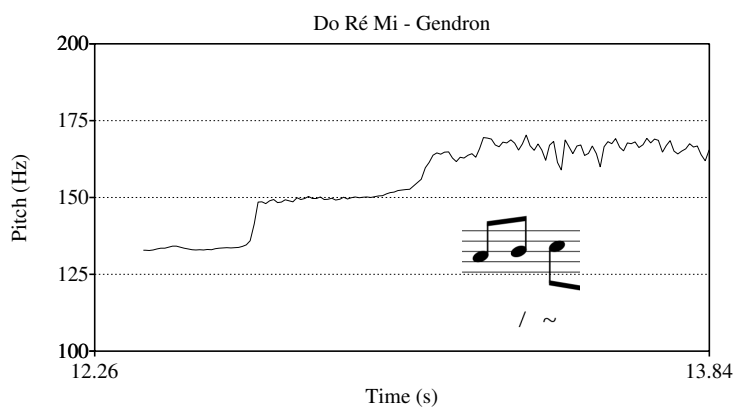
Figure 3: Fundamental frequency contour (do ré mi) extracted from Gendron's recording, with the software Praat, with the corresponding annotated score excerpt. The portamendo goes from D to E.

| < | crescendo (for long notes, and the attack note) |
|---|---|
| > | decrescendo |
| + | loudest note of a three-quavers pattern |
| = | note which beat duration is longer than the one of its to direct neighbours notes |
| _ | note which is played longer than the mean beat duration of the whole phrase |
| ⌣ | vibrato |
| / or \ | portamento, from this note to the following one |

Figure 4: Dictionary of annotation symbols.

the score. The score is written as an .xml file, and I used the software MuseScore to display the music and annotations.

## 3.1 Models implementation

I implemented the Implication-Realisation model (see 1.3.1), as well as two components from the tonal tension model: the local dissonance and the melodic attraction (see 1.3.2).

### 3.1.1 Melodic attraction

The melodic attraction considers a pair of two notes. The numerical calculation of the factor of attraction of the first note by the second one considers the distance between the two, and their respective stability (or anchoring strength) within a key [10]. The anchoring strength represents how stable a tone is in a tonality. In C major, the C is the most stable note: its anchoring strength is $s_C = 4$. It is followed by G and E ($s_E = s_G = 3$) because the fifth and the third build the scale. Then come the other notes of the diatonic scale (D, F, A, B: $s = 2$), and non-harmonic tones are the less stable tones (for C#, Eb, F#, G#, Bb: $s = 1$) (Fig. 13). The contribution of the distance between the two notes (number of semitones that separate them) is an imitation of Newton's law of gravitational attraction:

$$\alpha_{n_2 \to n_1} = \frac{s_2}{s_1} \cdot \frac{1}{d^2}$$

where $\alpha$ is the attraction of note 1 ($n_1$) by note 2 ($n_2$), $s_1$ and $s_2$ are the anchoring factors of the two notes, and $d$ is the distance in semitones between $n_1$ and $n_2$.

In this definition, the melodic attraction of the leading tone by the tonic is very high, as the anchoring strength of the first is low and the one of the latter is very high, and they are separate by a very small distance (Fig. 13).

The code calculates, for each pair of notes, their anchoring strength and the attraction factor. It runs as a loop over the list of notes, and returns the attraction factor for each note of any .xml melodic score.

Important is to note that the melodic attraction needs a basic space (the key) to evaluate the anchoring strength of the tones in that key. In my model, the basic space is D minor and is not able to take into account any modulation. As we focus on one theme, it is relevant to stay in D minor all along.

### 3.1.2 Local dissonance

The surface tension, or local dissonance, looks at every vertical event: every instant a note is played starts a new event, and it contains all the notes that are played at this time (even if they don't begin at the same time). It is an additive combination of the parameters described in the tonal tension model [10]: the note which is in the upper voice (if it is the tonic, it is more stable than if we have the third or the fifth), the chord inversion (it is considered to be more dissonant

**Surface tension rule**: $T_{diss}(\mathbf{y})$ = scale degree + inversion + non-harmonic tones (summed over all the pitch classes in y's span), where

scale degree = 1 if $\hat{3}$ or $\hat{5}$ in the melodic voice, 0 otherwise;

inversion = 2 if $\hat{3}$ or $\hat{5}$ in the bass, 0 otherwise;
non-harmonic tone = 3 if a pitch class is a diatonic non-chord tone, 4 if it is a chromatic non-chord tone, 0 otherwise.

Figure 5: Rules of calculating the local dissonance, from [10].

if the chord is not in fundamental state), and the presence of non-harmonic tones (Fig. 5). It runs as a loop over the score, and returns the degree of dissonance for each event[3].

### 3.1.3  Implication-Realisation Model

The Implication-Realisation model considers a set of three notes. Considering the sizes and directions of the two consecutive intervals built by these notes, an expectation score is assigned to the third note. It results from the additive contribution of five criterion [14].

Though, the IR model with its five principles has been shown to be overspecified and Schellenberg proposes a simplified model [15] [17]. He showed that the five criterion were redundant, and built a three-principles parsimonious model. I finally used this simplified model, which gives similar results to the original one. These principles are:

- Registral Direction: a large interval implies a change in direction. If the first interval is large (more than six semitones), the second one is more expected to change direction. If we have a large first interval without a change in direction, the registral direction is negative $(-1)$. If the first interval is small, we don't build expectation about the direction of the following one.

- Registral Return: the third tone is likely to return close to the first one. If it falls within two semitones from the first tone, it is coded as 1, 0 otherwise.

- Proximity: it is coded 0 if the second interval's size is 0 semitones, -1 if it is 1, -2 if 2 and so on. It accounts for the likelihood of the third tone to stay close to the preceding one.

I implemented the automatic calculation of this model. For each note of a melodic score, regarding the two preceding ones, it calculates the three parameters and adds them.

It is, apparently, the first time it has been applied to a sonata.

---

[3]From the Tonal Tension model remains the distance between chords, that I have not implemented. The authors of the model explain that the more relevant way to put events in relation is the sequential way. This needs a score-reduction, and this analysis operation is not completely systematic. Moreover, it does not reflect the perception of all listeners: depending on ones own musical experience and knowledge of the piece, one listener may hear events sequentially or more hierarchically, putting events in relation at a larger scale. As Debussy's sonata is not written in a tonal way, especially in our excerpts, it will not make any sense to look at a distance between chords, that were not thought as chords.
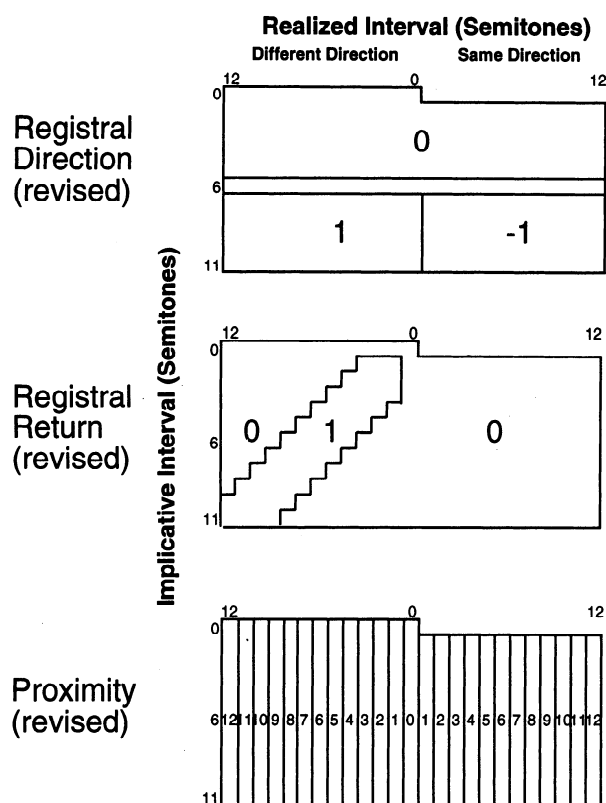
Figure 6: Graphical description of the revised I-R model. Vertically: the size (in semi tones) of the first interval; horizontally: the (algebraic) size of the second one. The three tables give scores that are then added together. The coefficient named proximity has to be taken negative. *Schellenberg 1996* [17].
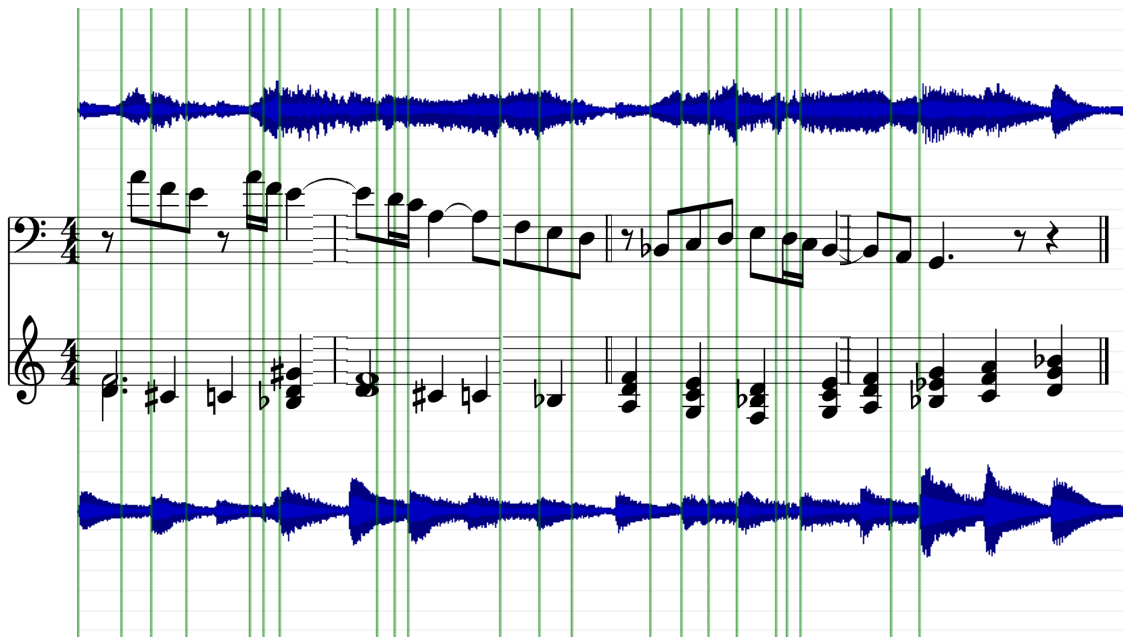
Figure 7: Wave function of Tortelier's recording, accompanied by the corresponding score. Vertical bars are the annotated onsets of the cello's notes.

## 3.2 Recordings analysis and annotation process

### 3.2.1 Timing

Before beginning to compute or to build any model, we need to extract the onsets of the played notes from the recording. Using the software Sonic Visualiser, I annotated manually the onsets times of the notes played by the cellist (Fig. 7). From these data, we can calculate the note's duration, and start to do some comparisons.

### 3.2.2 Local beat duration

I wrote a program in Python to calculate the local beat duration $D(n)$ for every note $n$. We define it as the ratio of duration of the note as it is played divided by its score duration[4] (the quarter length qL is the unit used to measure notes' durations in the library music21. It corresponds to the number of seconds per quarter note ($\quarternote$). The duration of one quarter note is $T(\quarternote) = 1$ qL, for a eight: $T(\eighthnote) = 0.5$ qL, and so on.). For a note $n$:

$$D(n) = \frac{t_{n+1} - t_n}{T(n)}$$

---

[4]It is proportional to the inverse of a tempo. The tempo in bpm (beat per minute) is $\frac{60}{D}$.

where $t_n$ are the onset times in seconds, $T(n)$ is the note's score-duration in qL. $D(n)$ is then in s.qL$^{-1}$ .

We define then the mean local beat duration for one phrase of $N$ notes:

$$D_m = \frac{\sum\limits_{n=1}^{n=N} D(n)}{N}$$

If $D(n) > D_m$, the note $n$ (played longer than the mean tempo) gets the annotation ◡. It lets us easily see where the descelerando are. The annotation symbols are written as lyrics, an attribute of the objects notes, rests and chords.

If $D_n > D_{n-1}$ and $D_n > D_{n+1}$, the note $n$ (played longer than its two neighbours notes) gets the annotation =. It is a more local observation, which is meaningful for the stress of notes at a local scale.

### 3.2.3   Loudness

For the loudness calculation, I used the MA (Music Analysis) toolbox, written in Matlab [18]. It uses a psychoacoustic model for evaluating the perceived loudness. From the audio file, it returns the loudness of the signal in sones. The sone is a unit for measuring perceived loudness, developed by psychoacoutsics researchers [19]. One sone corresponds to the loudness level of a pure 1 kHz signal[5] at 40 dB. It is a linear scale: 2 sones will be the double perceived loudness[6].

The sone calculation is based on a psychoacoustic model of loudness perception [20]. It takes into account that our perception of loudness is not the same for all frequencies, and also the masking effect of the ear[7]. The sound is then decomposed into 24 frequencies band, for which the loudness sensation is calculated. They are then brought together to output the global loudness [21].

### 3.2.4   Notes extraction

I wrote a Python code to cut and extract, from the audio file, the notes whose loudness I want to calculate. Thanks to the onset annotation file, I can select the time range and extract the note automatically. After having extracted these notes, I use the Matlab code to run the sone calculation. Then, I come back to Python for the annotation process.

I extracted the notes that are interesting to look at in term of loudness (see 2.2). First, the longest notes (dotted crochets) and the attack notes were cut off from the audio files. Then we calculate the slope of the loudness. If the slope is positive and higher than a threshold of 0.01, the note will be Being aware of the fact that, for each dotted crochet, the piano plays a

---

[5]Stevens' definition of sones [19].

[6]To have an idea, a leaves's noise or calm breathing is around 0.02 sones, the threshold of pain is around 600 sones. A normal conversation goes from 1 to 4 sones. In our sonata, sones can reach up to 15 sone

[7]In some situations, one sound otherwise clearly audible can be masked by the presence of another one. It is the case for a conversation in the street when a loud truck drives past, but it also can happen in music.

chord around the second third of the note duration, I decided to remove the highest value of the loudness derivative. It corresponds to the step in loudness we can observe at every piano onset. Nevertheless, we have to keep in mind that the loudness we calculate is never the cello loudness only, as we do not separate the two instruments. But assuming that the decrease in loudness of the piano notes will always be the same. And considering that the important factor is what the listener can perceive, so the total loudness, is due to both instruments, it does sense to calculate the loudness of the cello and the piano together.

Then, the three quavers from the different patterns were automatically extracted as well (Fig. 1), before running the sone calculation for each of them. Then, back to Python, the mean loudnesses of the three notes of one pattern are compared together, and we annotate the loudest one with a +.

I did the detection of vibrato and portameto manually: listening to the recordings, I marked up for each note if it is played with or without vibrato. The same for portamento (they are very few in these excerpts).

## 4   Analysis results

All the annotated scores for all cellists of my data set are available in the appendix D. We will discuss here some interesting observations we can draw from the results.

### 4.1   Long notes

From the annotated scores, we can observe that in our recordings, all cellists have played the dotted crochets decrescendo, apart from Piatigorsky. He is the only one to play the long notes with crescendo. That makes an individual gesture of Piatigorsky's playing (Fig. 8).

### 4.2   Patterns annotation

The three-quavers patterns annotations show which one of the three notes is played louder than the other two (Fig. 1). We can observe two different correlations of the pattern shaping with the melodic attraction and the implication-realisation model: for Tortelier, Gendron and Piatigorky, F is the loudest note (Fig. 10), and is, within this pattern, the note with the highest attraction factor toward the following E. Moreover, it is coherent with the second occurrence of the theme in the recapitulation: they play louder the C than Bb and D, and this is the note with the higher attraction factor. The cellists enhance here the note which has the highest factor of melodic, that is to say, the note that is the most attracted by the following one.

On the other hand, Maisky and Beer shape the motive in another way: they play louder the E, in the first half of the theme, and the Bb in the second half (Fig. 11). And again, they are consistent (although we can notice that in Beer's, the F gets the annotation +, instead of the E. It is due to the wolf note, that is very audible on Beer's cello. Because of that, the F is heard louder).
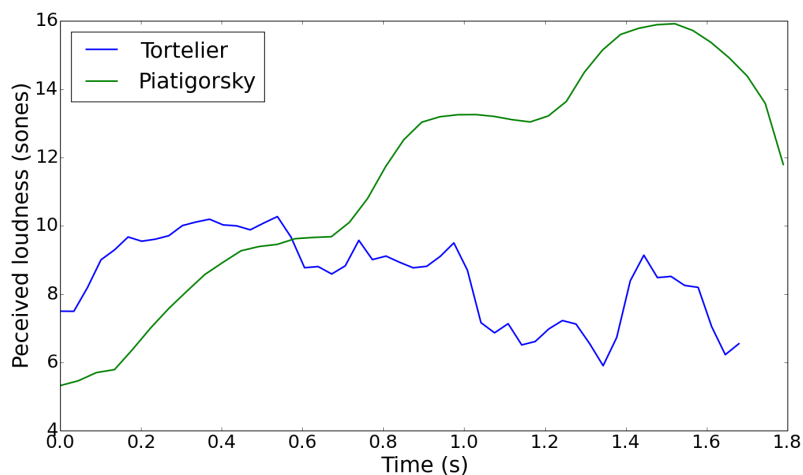
Figure 8: Perceived loudness (in sones) of the dotted crochet 'Mi' for Tortelier and Piatigorsky's recordings. We can observe Piatigorsky's crescendo.
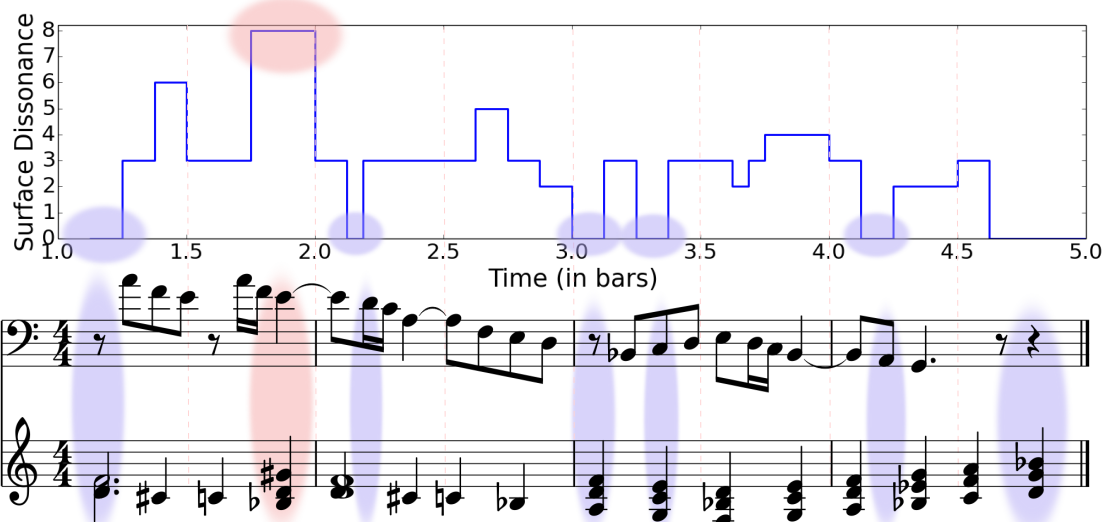


Figure 9: Graph of the surface dissonance (from the tonal tension model) for the first half of the first theme. In blue: no dissonance; in red: the highest dissonance level in this phrase.

Looking at timing (Appendix D), there is no systematic way of rhythmically play the three-quavers patterns: cellists' playing do not concur on a common rhythmic accentuation. Every one has his own way to variate timings.
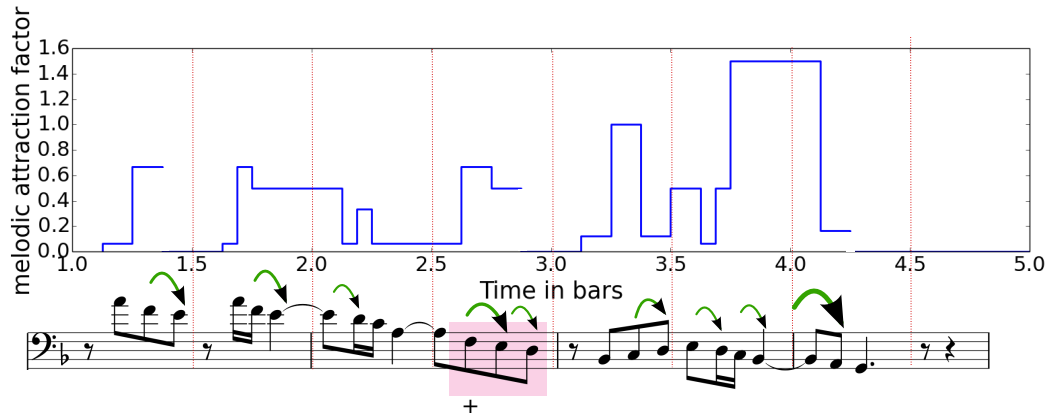


(a)
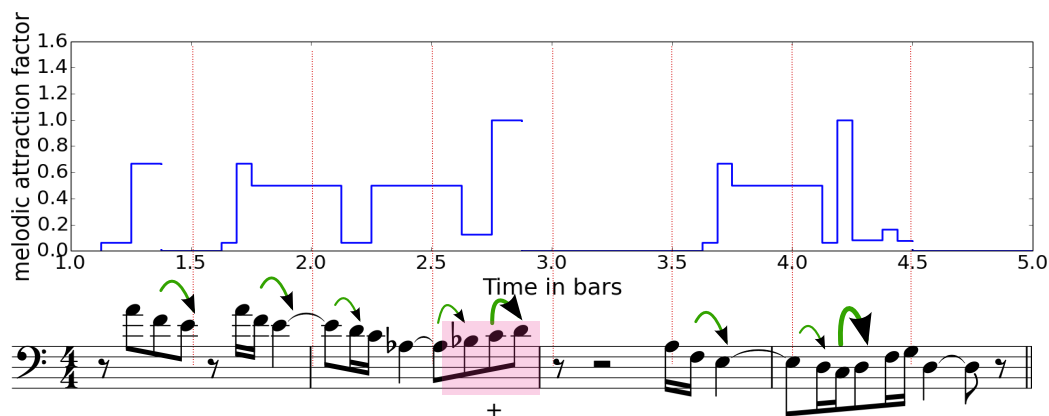


(b)

Figure 10: Melodic attraction score (from the tonal tension model) with the three-quavers pattern annotation, for the cellists Tortelier, Piatigorsky and Gendron. (a) first part of the exposition theme ; (b) second part.

## 4.3   Attack of the first note

Almost all cellists play the first note with a crescendo, except Tortelier, but not for each phrase. We can hear that for one of the phrases (beginning of the theme in the recapitulation), the attack of the first note is played harder.

(a)



(b)
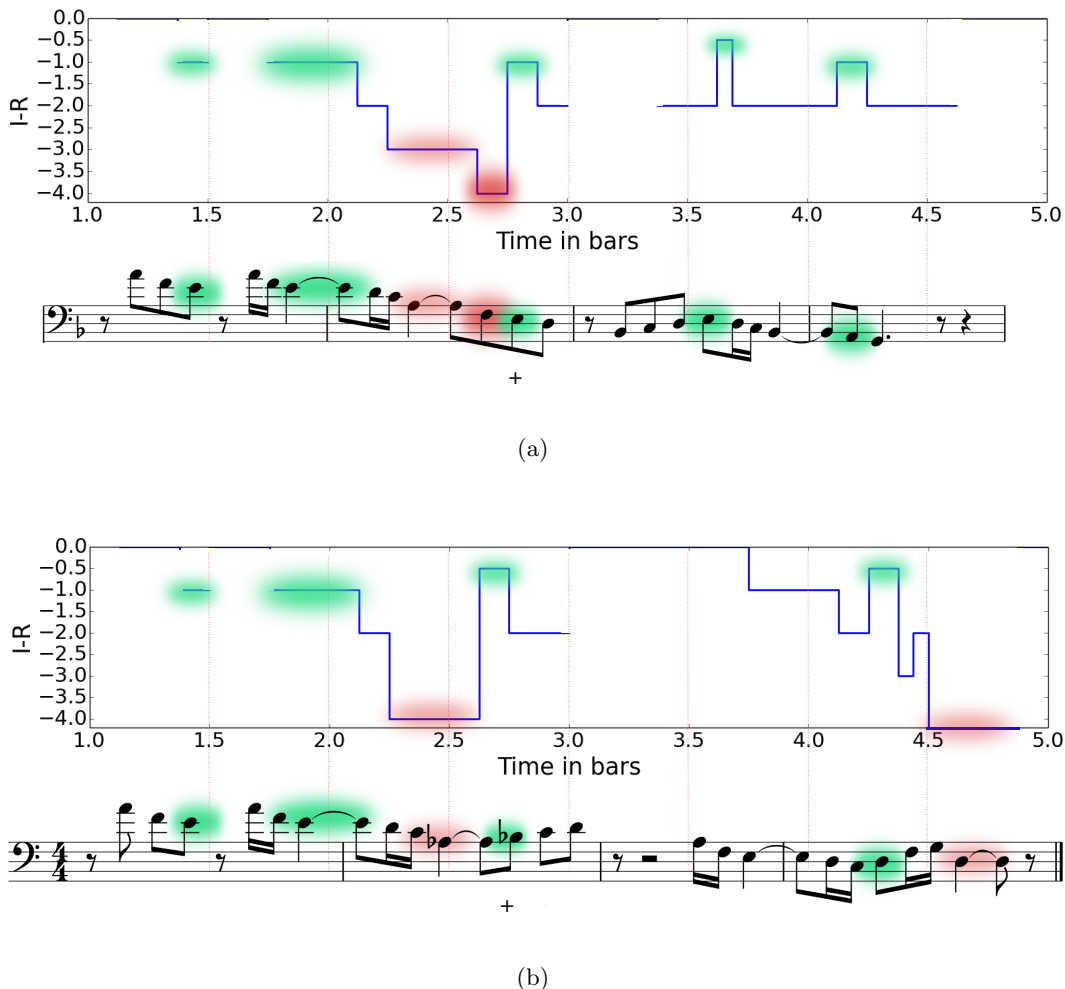
Figure 11: Expectation score (Implication-Realisation model) with the three-quavers pattern annotation, for the cellists Maisky and Beer. In green: notes that have a high expectation score (they are considered attended); in red: notes that have a low expectation score. (a) first part of the exposition theme ; (b) second part. They both emphasize the most expected note.

## 4.4 Timing annotations

We can easily observe that the first theme ends with a *rallentando* (slow down) (see D), by the annotated notes (♩) on the end: all these notes are played longer than the mean beat duration of the excerpt. In the recapitulation, the excerpt is followed by another phrase (the cello plays the piano opening theme of the sonata). That is why there is no very obvious *rallentando*, but we can still note that most of the notes in the second half of the excerpt are played slower (see D).

A common expressive feature in timing is to be observed in the third bar of the first half of the theme, in the exposition as well as in the recapitulation (see Appendinx D). The highest note of the motive E is played longer than its neighbour notes (= ♩), after that the notes Bb C D are played faster (no annotation symbol below them). This feature is shared among all

17

cellists, except for cellist Beer, who has a very smooth and regular playing.

## Conclusion

I have implemented the automatic calculation of the implication-realisation model, as well as the melodic attraction and the local dissonance from the tonal tension model, which can run on any score, coded as an .xml file. Even if we have to keep in mind that the the tonal tension model relates to tonal music, it is worthy to apply it to some passages of Debussy's sonata. Indeed, we could find some interesting behaviours in the cellists' playings that can be related to the two models' predictions.

The melodic attraction is calculated according to a key. The piano part does not obviously refers to D minor. The cello melody can still be heard as D minor or Dorian. Even if the reference to tonality is not clear, the melodic attraction is able to give some information. Concerning the local dissonance calculation, it is also related to a tonal use of music. The consideration of dissonances by Debussy is not tonal. But the model is built on the human perception of dissonance in western tonal music. Our ear is mainly trained to this type of music, so finally it is not completely irrelevant to give some credit to this model.

As the Implication-Realisation model builds expectations values on three-notes patterns, it does not take into account the memory of the listener, which plays an important role in the listening experience: our anticipation in melody is strongly influenced by the hearing of patterns that we expect to be repeated, with or without transposition and variations. In the opposite direction, we are able to anticipate a melody at a lager scale than one note, as well as we can remember more than two tones. With its small-scale considerations, the model just pretends to quantize note to note expectation, and it is enough to observe some correlations with interpretation features.

The annotation process I have initiated can be a start for a systematic annotation system of musical expressiveness. In our particular case of study, it is tightly related to string instruments, and to the particular theme I chose to analyse. The parameters I extracted are indeed very dependant on the piece we choose to analyse. For instance, the three-quavers patterns and the dynamics of dotted crochets are interesting in this theme, but another work or another passage of the same piece will call for other detailed annotations. A larger dictionary may have to be built, as well as a system for systematic pattern recognition.

This annotation system can be particularly useful for cognitive scientists who look at emotion and expressivity in music.

A bigger data base would be needed, and so as to build it, an automatic onset detection would be necessary. The manual onset annotation requires a lot of time to have precise data. Some onset detection programs are already available for piano solo. However, the cello playing allows very smooth attacks and releases, as well as the overlap of notes, especially when they are played on different strings. That makes automatic detection highly challenging. In some cases, it is more a matter of perception, than a matter of effective playing.

The parameters extracted here are low level features. So as to reach a higher level annotation scheme, we could start annotating the stress of notes: where notes are played longer and louder for instance. It would be another direction to continue this work. From the set of symbols developed here, we can already see some patterns of stress. The idea is to mark up directions, stress and emphasis in phrasing.

## Acknowledgements

## References

[1] Avo Somer. Musical syntax in the sonatas of debussy: Phrase structure and formal function. *Music Theory Spectrum*, 27(1):67–95, 2005.

[2] Fred Lerdahl. Calculating tonal tension. *Music Perception*, pages 319–363, 1996.

[3] Eugene Narmour. *The analysis and cognition of melodic complexity: The implication-realization model.* University of Chicago Press, 1992.

[4] Mary E Beckman and Julia Hirschberg. The tobi annotation conventions. *Ohio State University*, 1994.

[5] Kim EA Silverman, Mary E Beckman, John F Pitrelli, Mari Ostendorf, Colin W Wightman, Patti Price, Janet B Pierrehumbert, and Julia Hirschberg. Tobi: a standard for labeling english prosody. In *The Second International Conference on Spoken Language Processing, ICSLP 1992, Banff, Alberta, Canada, October 13-16, 1992*, 1992.

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

[7] Ching-Hua Chuan and Elaine Chew. A dynamic programming approach to the extraction of phrase boundaries from tempo variations in expressive performances. In *ISMIR*, pages 305–308, 2007.

---

[8]You are very French, you should play Debussy's Sonata!

[8] Ju-Lee Hong. Investigating expressive timing and dynamics in recorded cello performances. *Psychology of Music*, 31(3):340–352, 2003.

[9] Christopher Raphael. Representation and synthesis of melodic expression. In *IJCAI*, pages 1475–1480, 2009.

[10] Fred Lerdahl and Carol L Krumhansl. Modeling tonal tension. 2007.

[11] Fred Lerdahl and Ray Jackendoff. *A generative theory of tonal music*. MIT press, 1985.

[12] Caroline Palmer. Anatomy of a performance: Sources of musical expression. *Music Perception*, pages 433–453, 1996.

[13] M Farbood. A global model of musical tension. *Proceedings of the 10th International Coference on Music Cognition and Perception*, 2008.

[14] Carol L Krumhansl. Music psychology and music theory: Problems and prospects. *Music Theory Spectrum*, 17(1):53–80, 1995.

[15] E Glenn Schellenberg. Expectancy in melody: Tests of the implication-realization model. *Cognition*, 58(1):75–125, 1996.

[16] Lola L Cuddy and Carole A Lunney. Expectancies generated by melodic intervals: Perceptual judgments of melodic continuity. *Perception & Psychophysics*, 57(4):451–462, 1995.

[17] E. Glenn Schellenberg. Simplifying the implication-realization model of melodic expectancy. *Music Perception: An Interdisciplinary Journal*, 14(3):pp. 295–318, 1997.

[18] Elias Pampalk. A matlab toolbox to compute music similarity from audio. In *ISMIR*, 2004.

[19] Stanley Smith Stevens. A scale for the measurement of a psychological magnitude: loudness. *Psychological Review*, 43(5):405, 1936.

[20] E Zwicker and H Fast. Psychoacoustics. 1999.

[21] William M. Hartmann. *Signals, sound, and sensation*. Springer Science & Business Media, 1997.

# A    Scores of the excerpts



(a)



(b)



(c)



(d)

Figure 12: Score of the excerpts of the sonata. (a) and (b) are the theme in the exposition, (c) and (d) are the second presentation of the theme in the recapitulation.

# B  Anchoring

(a)

| Anchoring strength | The basic space with the fifths level omitted |
|---|---|
| 4 | 0 |
| 3 | 0     4     7 |
| 2 | 0   2   4 5   7   9     11 |
| 1 | 0 1 2 3 4 5 6 7 8 9 10 11 |

(b)

**Melodic attraction**: $\alpha(p_1 \rightarrow p_2) = \frac{s_2}{s_1} \times \frac{1}{n^2}$, where $p_1$ and $p_2$ are pitches, with $p_1 \neq p_2$; $\alpha(p_1 \rightarrow p_2)$ = the attraction of $p_1$ to $p_2$; $s_1$ = the anchoring strength of $p_1$ and $s_2$ = the anchoring strength of $p_2$ in the current configuration of the basic space; and $n$ = the number of semitone intervals between $p_1$ and $p_2$.

(c)

$$\alpha(B \rightarrow C) = \frac{4}{2} \times \frac{1}{1}^2 = \frac{4}{2} = 2$$
$$\alpha(D \rightarrow C) = \frac{4}{2} \times \frac{1}{2}^2 = \frac{4}{8} = 0.5$$
$$\alpha(F \rightarrow E) = \frac{3}{2} \times \frac{1}{1}^2 = \frac{3}{2} = 1.5$$
$$\alpha(E \rightarrow F) = \frac{2}{3} \times \frac{1}{1}^2 = \frac{2}{3} = 0.67$$

Figure 13: Description of the melodic attraction factor. *Lerdahl, Krumhansl 2007* [10]

# C  Score and Recordings

Score:

- *Sonate pour Violoncelle et Piano*, Claude Debussy, edition Durand & Cie., 1915.

Recordings used for the analysis:

- Paul Tortelier, Jean Hubeau - Label: Elatus - Released: 17/05/2006 - Recorded: 1962.

- Maurice Gendron, Jean Français - Label: Philips - Released: 12/1994.

- Gregor Piatigorsky, Lukas Foss - Label: RCA Victor Red Seal - Released: 1959.

- Mischa Maisky, Martha Argerisch - Label: His Master's Voice - Released: 1983 - Recorded: 12/1981.

- Susanne Beer, Gareth Hancock - Label: Divine Art - Released: 12/09/2008.

# D  Annotated scores

This appendix shows all the annotated scores for all cellist, all phrases. See the dictionary of annotation symbols in Fig. 4.

## First phrase, first half



Beer



Gend



Piat



Mais



Tort

## First phrase, second half



Beer



Gend



Piat



Mais



Tort

## Second phrase, first half



Beer



Gend



Piat



Mais



Tort

## Second phrase, second half