Prolegomena to Music Semantics*

Philippe Schlenker
(Institut Jean-Nicod, CNRS; New York University)

March 29, 2016

Preliminary draft - Version 1.0

Music consultant: Arthur Bonetto

Note: Whenever possible, links to audiovisual examples have been included in the text.

Abstract. We argue that a formal semantics for music can be developed, although it will be based on very different principles from linguistic semantics and will yield less precise inferences. Our framework has the following tenets: (i) Music cognition is continuous with normal auditory cognition. (ii) In both cases, the semantic content of an auditory percept can be identified with the *set of inferences it licenses on its causal sources*, analyzed in appropriately abstract ways (e.g. as 'voices' in some Western music). (iii) What is special about music semantics is that it aggregates inferences based on normal auditory cognition with further inferences drawn on the basis of the behavior of voices in tonal pitch space (through more or less stable positions, for instance). (iv) This makes it possible to define an inferential semantics but also a truth-conditional semantics for music. In particular, a voice undergoing a musical movement *m* is true of an object undergoing a series of events *e* just in case there is a certain structure-preserving map between *m* and *e*. (v) Aspects of musical syntax (notably Lerdahl and Jackendoff's 'time-span reductions') are derivable on semantic grounds from an event mereology ('partology'), which also explains some cases in which tree structures are inadequate (overlap, ellipsis). (vi) Intentions and emotions may be attributed at three levels (the source, the musical narrator, the musician), and we speculate on possible explanations of the special relation between music and emotions. Finally, (vii) we argue that two empirical methods may prove useful to study music semantics: in special cases, one may *decompose* a piece into its component parts (e.g. rhythm, melody) to assess their individual semantic effects; in the general cases, one may rewrite part of a piece (e.g. with changes of harmony) in order to obtain *minimal pairs* whose semantic effects can be contrastively assessed.

* As the title indicates, these are preliminary remarks on a complex topic.

Arthur Bonetto has served as a regular and very insightful music consultant for these investigations; virtually all musical examples were discussed with him, and he played a key role in the construction of all minimal pairs, especially when a piece had to be rewritten with special harmonic constraints. However he bears no responsibility for theoretical claims – and possible errors – contained in this piece.

For helpful conversations, many thanks to John Bailyn, Karol Beffa, Arthur Bonetto, Laurent Bonnasse-Gahot, Emmanuel Chemla, Didier Demolin, Paul Egré, Ray Jackendoff, Jonah Katz, Fred Lerdahl, Salvador Mascarenhas, Rob Pasternak, Claire Pelofi, Martin Rohrmeier, Benjamin Spector, Morton Subotnick, Francis Wolff, as well as audiences at New York University and SUNY Long Island. I learned much from initial conversations with Morton Subotnick before this project was conceived. Jonah Katz's presence in Paris a few years ago, and continued conversations with him, were very helpful. I have also benefited from Emmanuel Chemla's insightful comments on many aspects of this project. None of these colleagues is responsible for any errors made here.

The research leading to these results received funding from the European Research Council under the European Union's Seventh Framework Programme (FP/2007-2013) / ERC Grant Agreement N°324115–FRONTSEM (PI: Schlenker). Research was conducted at Institut d'Etudes Cognitives, Ecole Normale Supérieure - PSL Research University. Institut d'Etudes Cognitives is supported by grants ANR-10-LABX-0087 IEC et ANR-10-IDEX-0001-02 PSL*.

1 Introduction

While the *syntax* of music has been studied in great formal detail (see Jackendoff and Lerdahl 1983, Lerdahl 2001 for classical music), the topic of music *semantics* has not given rise to the same formal developments. One possible reason is that 'music semantics' has no subject matter: while the existence of rules that constrain musical form is not in doubt, there might be no such thing as a *semantics* of music. By 'semantics', we mean *a rule-governed relation between the musical form and some music-external reality*, no matter how abstract. The 'no semantics' position might well be the Null Hypothesis: there is little initial reason to think that music is endowed with either truth conditions or denotations. By contrast, speakers of a language have no trouble deciding under what conditions a well-formed sentence is true, which has motivated the development of a sophisticated truth-conditional semantics in contemporary linguistics. In music, in most cases one would have considerable trouble putting in words what the music conveys, besides vague and impoverished descriptions that often have to do with the emotions that a piece may evoke in the listener.

Despite these initial qualms, we explore the view that music has a semantics, albeit a very different one from natural language: first, music semantics usually conveys much more abstract information than language does; second, and more importantly, its informational content is derived by very different means. Our guiding intuition is that *the meaning of a musical piece is given by the inferences that one can draw about its virtual sources*[1], which in salient cases can be identified with the 'voices' of classical music theory. Our analysis is in two steps.

• First, we take properties of normal (non-musical) auditory cognition to make it possible to identify one or several virtual sources of the music, and to license some inferences about them depending on some of their non-tonal properties (rhythm, loudness, patterns of repetition, etc). Thus *music semantics starts out as noise semantics*. Importantly, these sources are fictional, and need not correspond to actual sources: a single pianist may play several voices at once; and a symphonic orchestra may at some point play a single voice.

• Second, we take further inferences to be drawn about these sources from their behavior within tonal pitch space. This space has non-standard properties, with different subspaces (major, minor, with different keys within each category), and locations (chords) that are subject to various degrees of stability and attraction. Inferences may be drawn on a (virtual) source depending on its behavior in that space.

A metaphor might clarify what we have in mind. If we see various objects on a roller-coaster, we may draw all sorts of inferences about them from their behavior and especially their movement – for instance, their movement may or may not seem to be goal-directed. But these inferences will crucially depend on the interaction between these objects and the physical properties of the roller-coaster: if an object moves against gravity in a particularly steep area, we can infer that it might be self-propelled; if it remains in a point of great instability, we may expect that it will fall to a point of greater stability, etc. Similarly, in the musical space the inferences we draw on the sources of the music will depend on all sorts of properties of the tonal pitch space by which they are constrained.

We take the present the framework to integrate two intuitions that were developed in earlier analyses.

• In Bregman's application of Auditory Scene Analysis to music, the listener analyzes the music as a kind of 'chimeric sound' which 'does not belong to any single environmental object' (Bregman 1994 chapter 5). As Bregman puts it, 'in order to create a virtual source, music manipulates the factors that control the formation of sequential and simultaneous streams'. Importantly, 'the virtual source in music plays the same perceptual role as our perception of a real source does in natural environments'. This allows the listener to draw inferences about the virtual sources of the music: 'transformations in loudness, timbre, and other acoustic properties may allow the listener to conclude that the maker of a sound is drawing nearer, becoming weaker or more aggressive, or changing in other ways', although this presupposes an analysis in which these sounds are taken to reflect the behavior of a single virtual source.

• The other antecedent idea is that the semantic content of a musical piece is a kind of 'journey through tonal pitch space'. Lerdahl 2001 thus analyzes 'musical narrativity' connection with a linguistic theory (Jackendoff 1982) in which 'verbs and prepositions specify places in relation to starting, intermediate, and terminating objects'. For him, music is equally 'implicated in space and

[1] The term 'virtual source' is due to Bregman, e.g. Bregman 1994.

motion': 'pitches and chords have locations in pitch space. They can remain stationary, move to other pitches or chords that are closer or far, or take a path above, below, through, or around other musical objects'. More recently, Ganroth-Wilding and Steedman 2014 provide an explicit semantics for jazz sequences in terms of motion in tonal pitch space.

It is essential for us that that these two ideas should be combined within a single framework. An analysis based on Auditory Scene Analysis alone might go far in identifying the virtual sources and explaining some inferences they trigger on the basis of normal auditory cognition, but it would fail to account for the further inferences that one draws by observing the movement of the voices in tonal pitch space – for instance the fact that a dissonance yields an impression of instability, while a tonic chord may give an impression of repose; or the fact that the end of piece is typically signaled by a movement towards greater tonal stability. Conversely, an analysis based solely on motion through tonal pitch space would miss many of the inferences about the sources that are drawn on the basis of normal auditory cognition. For instance, some classical pieces end with a gradually decreasing speed and volume. While these could be taken to be conventional ways to mark the end of a piece, it is plausible that they signal that the virtual source of the music is gradually losing energy to eventually die out. This inference crucially builds on properties of normal auditory cognition in the non-musical world, and it is essential to take these into account to have a full account of inferential effects in music.[2]

While we will informally develop our analysis in inferential terms, by collecting the diverse inferences triggered by a musical piece (some of them based on normal auditory cognition, others on properties of tonal pitch space), we will provide and exemplify a notion of musical truth. In a nutshell, a voice undergoing a musical movement *m* is true of an object undergoing a series of events *e* just in case there is a certain structure-preserving map between *m* and *e*. Somewhat similarly, a visual animation can be taken to be true of a sequence of events just in case the events resemble the animation in appropriate ways, preserving certain geometric rather than auditory properties. In most case, the informational content of a musical piece will be far more abstract than information conveyed by natural language sentences; more importantly, this informational content will be derived by entirely different means (a source-based semantics rather than a compositional semantics).

The rest of this article is organized as follows. In Section 2, we sketch what we take to be the Null Hypothesis: music has a syntax and possibly a pragmatics, but not semantics. It thus takes a detailed empirical argument to show that a semantic approach is legitimate. In Section 3, we provide preliminary examples of semantic effects in a musical piece. In Section 4, we list systematic effects that are derived from normal auditory cognition. In Section 5, we list further semantic effects that are drawn on the basis of tonal properties. We sketch an analysis that integrates both types of inferences in Section 6, with a very simple 'toy model' that illustrates our approach to 'musical truth'. We then argue in Section 7 that a semantic approach makes it possible to revisit certain aspects of musical syntax (Lerdahl and Jackendoff's 'grouping structure' and 'time-span reductions'), and to explain why tree structures are often useful, but are sometimes overly constrained. In Section 8, we explore various levels of pragmatic analysis in music, before speculating in Section 9 on how our framework could account for the special role that emotions play in music. We sketch various extensions and comparisons in Section 10, a draw some conclusions in Section 11.

2 Music without meaning: the Null Hypothesis

The view that one can define a 'music semantics' is controversial, and should be argued for on detailed empirical grounds. We start by articulating what we take to be a null hypothesis according to which *music has as a syntax and a pragmatics but crucially no semantics*. We do so for two reasons. First, this is certainly the simplest view, and it is important to see how far it can take us in the analysis of musical effects. Second, by highlighting the properties that can be captured *without* a semantics, we will be in a better position to assess the special role of semantics proper, as well as the distinction between semantics and pragmatics. Later sections will provide examples of genuine semantic effects in music, and they will sketch a framework in which these can be captured.

2 In terms of Peirce's tripartition between *icons, indexes and symbols*, musical pieces should be viewed in a source-based semantics as indexes because they are representations "whose relation to their objects consists in a representation in fact" (Atkin 2010, Peirce 1868). In particular, in the general case the present analysis will not take musical pieces to be icons, which would involve a 'likeness' between the representations and their objects. However in cases in which an inference is drawn on the basis of non-musical auditory cognition, a musical piece will come to signify by *resembling* auditory aspects of its source; in such cases, some inferences will be iconic in nature (and the some musical sounds may be termed 'indexes' in Peirce's terminology). See for instance Koeulsch 2011 for a discussion of Peirce's tripartition in a musical context.

2.1 Musical syntax

It is probably uncontroversial that music has a syntax, defined as a set of principles that govern the well-formedness of musical pieces. We need not take a stand as to whether well-formedness is categorical or gradient. Nor do we need to take a stand on the formal properties that musical syntax has. A highly articulated view can be found in Lerdahl and Jackendoff's (1983) groundbreaking work on this topic (for a view that connects musical syntax more strongly to linguistic syntax, see Pesetsky and Katz's 2009).

For purposes of comparison with language, it will be useful to give ourselves a toy formal system that has a much simpler syntax. Its lexicon is made of three syllables, *la, lu, li*. A well-formed sequence is any sequence made of the sub-sequences *la lu* and *la li*. Everything else is ill-formed. Two possible ways of defining this very simple grammar are given in (1), and some examples are provided in (2). (The first grammar in (1)b makes use of the formal of 'context-free grammars', which are standardly assumed – with additional devices – in linguistics. The second grammar in (1)b makes use of the strictly less expressive formalism of 'regular grammars', which define finite-state languages.)

(1) a. Lexicon: Lex = {la, lu, li,}
b. Syntax
(i) Context-free grammar:
$S \rightarrow L, L\ S$
$L \rightarrow$ la lu, la li
(ii) Regular grammar:
$(\text{la lu} \cup \text{la li})^*$

(2) Examples
[la lu]
[la li]
[la lu] [la lu] [la lu] [la li] [la lu]

2.2 No semantics or an internal semantics

A natural view is that music simply has no semantics, and that it is a formal system that does not bear any relation akin to *reference* with anything. A slightly different view is that music has a semantics, but which pertains to objects that are themselves musical in nature – what we will call an 'internal' semantics. While these two views are distinct, they both differ from the view we will develop in this piece, according to which music has a natural semantics that establishes a relation between musical pieces and the music-external reality (see Meyer 1956 and Wolff 2015 for a broader discussion of this general debate).

Before we say a word about the 'internal' semantics in music, it might help us clarify our ideas to explore how such semantics can be constructed for a system as simple as the *la li lu* example of the preceding section. They key is that a syntactic system that has no semantics can still be *endowed* with a semantics that pertains to the *form* of the expressions themselves. In (3) we have done so for the context-free grammar defined in (1)b. Just as is standard for human language, each step in a derivation tree is interpreted by a semantic step. The result is not exciting: each syllable denotes itself, and each sequence denotes itself as well, with the proviso that the interpretation procedure adds pauses between groups of 2 syllables. Some simple examples are given in (4).

(3) a. Lexical semantics:
[[la]] = *la*
[[lu]] = *lu*
[[li]] = *li*

b. Compositional semantics
Notation: ^ is used to represent concatenation of expressions; for strings *s* and *s'*, *s*^_^*s'* denotes the concatenation of *s* and *s'* with a pause in between.

For any words *w*, *w'* of the lexicon Lex and for any sequences *l* and *s* of categories *L* and *S* respectively,
$[[\ [_L\ w\ w']\ ]] = [[w]]\hat{}[[w']]$
$[[\ l\ s\ ]] = [[l]]\hat{}\_\hat{}[[w']]$

(4) Examples
$[[\ [_L\ \text{la lu}]\ ]] = [[\text{la}]]\hat{}[[\text{lu}]] = \text{la}\hat{}\text{lu}$

[[[$_L$ la lu] [$_L$ la li]]] = [[[$_L$ la lu]]]^_^[[[$_L$ la li]]] = la^lu^_^la^li

Now this semantics adds very little to the syntax. But one could develop a more subtle variety of this internal semantics, one that only keeps track of certain properties of the form of our sequences. For example, in (5) we defines a semantics that keeps track of the vowels that appear at the end of our 2-syllable groups. Thus *la lu* will 'denote' *u*, while *la li* will 'denote' *i*, and the sequence *la lu la li* will denote the sequence *i^u*, i.e. the concatenation of the vowels *i* and *u*.

(5) Semantics based on vocalic paths
[[[$_L$ la lu]]] = u
[[[$_L$ la li]]] = i
For any sequences *l* and *s* of categories *L* and *S* respectively,
[[l s]] = [[l]]^[[w']]

(6) Examples
[[[$_L$ la lu]]] = u
[[[$_L$ la lu] [$_L$ la li]]] = [[[$_L$ la lu]]]^[[[$_L$ la li]]] = u^i

We can think of this semantics as associating with some strings a 'vocalic path' that tracks the sequence of some particularly important vowels that appear in it – here they are the non-predictable vowels of each 2-syllable group.

While no interesting analysis would postulate that music has the kind of semantics exemplified by (3), there are prominent examples of music semantics that develop more sophisticated versions of (5). Thus Granroth-Wilding and Steedman 2014 endow their formal syntax for jazz chord sequences with a semantics that encodes paths in a tonal pitch space whose structure is depicted in (7). In their analysis (framed within Combinatory Categorial Grammar), surface chords can be assigned syntactic categories that give rise to derivation trees. Each derivational step in the syntax goes hand in hand with a semantic step. And the semantics encodes movements in tonal pitch space. A minimal example is given in (8)

(7) Structure of the tonal pitch space assumed in Granroth-Wilding and Steedman 2014 (following Longuet-Higgins 1962a, b)

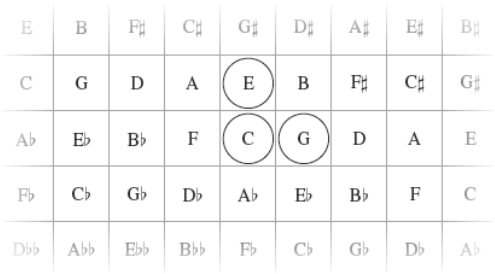


Figure 4: Part of the space of note-names (adapted from Longuet-Higgins, 1962a,b). Notes are separated by major thirds along the horizontal axis and perfect fifths along the vertical. The space extends infinitely in both dimensions. The circled points form a C major triad.

(8) Example of a syntactic and semantic derivation in Granroth-Wilding and Steedeman's (2014) framework (fragment of their Fig. 19)

$$\dfrac{\dfrac{V^7}{\lambda x.leftonto(x)} \quad \dfrac{I}{[\langle 0,0\rangle]}}{[leftonto(\langle 0,0\rangle)]}>$$

The final I denotes a location in tonal pitch space, with coordinates <0, 0>. The penultimate V^7 denotes a function from x to a position that ensures a 1-step leftward movement towards x – hence a movement from a G.

We believe that this analysis is close to an intuition developed in some of Lerdahl's work (2001), in which the meaning of music is essentially likened to a journey through tonal pitch space.

Importantly, this semantics is 'internal' – and thus not a 'real' semantics, from our perspective – because it does not draw a connection between music and the (music-)external reality, unlike the semantics we will argue for in this piece.

2.3 Pragmatics

Even if music has no semantics, it is natural to assume that it conveys information – namely about its own form. Thus music could in principle make use of certain devices to structure this information in optimal ways – one aspect of 'music pragmatics'.

To have a linguistic point of comparison, consider how language makes salient new elements. In (9)a, the second clause contrasts with the first in that *me* is replaced with *you*, and for this reason the new element *you* is focused (by way greater loudness, higher pitch and longer duration). If another element is focused instead, as is the case in (9)b, the result is deviant.

(9) a. He will introduce me to her, and then he will introduce YOU to her.
b. #He will introduce me to her, and then he will introduce you to HER.

These data could be analyzed in at least two ways. We could take the *form* of the elements – and specifically the fact that *you* is phonologically distinct from *me* – to be at the heart of the process. Alternatively, we could posit that what matters is that these two expressions have different *meanings*. Theories of linguistic focus are normally based on the second idea – for good reason: when expressions differ from each other only by their meaning, as in (10), one must still take into account their denotations to determine which ones count as 'new' (see Rooth 1996, Schwarzschild 1999, and also Arstein 2004 for some qualifications).

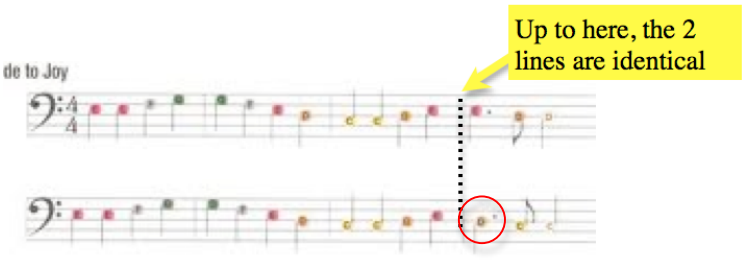
(10) a. [Talking about John]
I will introduce him to her, and then [turning towards Bill] I will introduce HIM to her.
b. Dialogue:
A: He will introduce me to her.
B. And then he will introduce ME to her.

Still, in systems that have no semantics (or only an internal semantics), form does play a role in contrastive focus assignment. Thus if I were to dictate to you a list of sequences produced by the *la li lu* grammar described above, I would certainly tend to focus (emphasize) elements that are new. For instance, in (11) it would seem natural to focus the syllable *li*, which contrasts with all the syllables encountered before, and in particular with all the 'parallel' syllables found at the end of the 2-syllable groups.

(11) [la lu] [la lu] [la LI] [la lu]

Do we find such effects in music? We might, as is illustrated in (12), where we feel that a performer might want to add greater emphasis on the first new note of the consequent (possibly realized by greater loudness and longer duration).[3] Importantly, there might other reasons why such an emphasis is found – in particular the fact that the note in question appears at the beginning of the cadence. Be that as it may, and whether emphasis reflects newness or something harmonic, it does appear to be used to structure musical information in appropriate ways. What matters for present purposes is that such pragmatic effects need not be indicative of a music semantics, since a formal system that has no semantics still conveys information about its own form (we come back to focus and information structure in Section 8.1)

(12) A focus accent in a music piece? (melody of Beethoven's *Ode to Joy*)



2.4 Summary and outlook

In this section, we made the following points:
(i) It is relatively uncontroversial that there is a musical syntax.

[3] To have a somewhat controlled minimal pair, one can try to play the entire passage in quarter notes, while putting an accent in the 'wrong' place, such as the 5th note (the second G) of the second line [https://soundcloud.com/philippeschlenker/beethoven-ode_to_joy]. Then one can artificially replace this overly accented G with its normal counterpart of the first line, while using a lowered version of the accented G to realize the circled D at the end [https://soundcloud.com/philippeschlenker/beethoven-ode_to_joy-d_focused].

(ii) Minimally, music conveys information about its own form. This can but need not be captured by defining a semantics in which music makes reference to music-internal properties. Still, this is not a semantics in the usual sense, as it does not connect music to a music-external reality.
(iii) It is plausible that music has a pragmatics, in the sense that one may modify the surface musical form (e.g. by way of musical accents) to highlight certain aspects of musical structure (old vs. new elements, or possibly expected vs. unexpected ones, less important vs. more important ones, etc). But this doesn't entail that music has a semantics in the usual sense – even meaningless strings of syllables are naturally produced with means (involving contrastive focus) that highlight aspects of their structure.

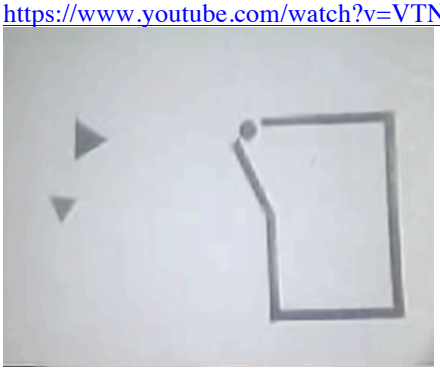
This Null Hypothesis is very plausible, and thus serious empirical arguments must be given to make plausible the project of a music semantics. We will do so in two steps: first, we will suggest that inferential properties of ordinary sound play a role in music; second, we will argue that further inferences are produced when we take into consideration specifically tonal properties of the music. We discuss examples that make the general program plausible in Section 3. Inferences derived from normal auditory cognition are discussed in Section 4, while inferences drawn from tonal properties are discussed in Section 5. While in these *Prolegomena* our arguments are entirely based on (hopefully shared) introspective judgments, we discuss at the end of each section the methods that could be used to establish experimentally the types of correlations we discuss (several of which were already studied with experimental means in the literature).

3 Examples of semantic effects

3.1 Visual examples

Since we wish to argue that an artificial system can trigger inferences about rather odd virtual sources of the percepts, it might be useful to start with a visual example that makes this point. Lerdahl 2001 makes reference to Heider rand Simmel's (1994) abstract animation "in which three dots moved so that they did not blindly follow physical laws, like balls on a billiard table, but seemed to interact with another – trying, helping, hindering, chasing – in ways that violated intuitive physics", and thus were perceived as animate agents. His point is to argue that similar effects arise in music: "here the dots are events, which behave like interacting agents that move and swerve in time and space, attracting and repelling, tensing and coming to rest". He concludes that "the remarkable expressive power of music is a manifestation of the internalized knowledge of objects, forces, and motion, refracted in the medium of pitches and rhythms."

(13) Stimulus from Heider and Simmel 1944
https://www.youtube.com/watch?v=VTNmLt7QX8E



In Heider and Simmel's animation (see (13)), the interpretation involves attributions of agency and intentions. But further and more basic properties can be attributed to abstract shapes as well. As an example, Kominsky et al. 2014 showed subjects abstract animations involving several pairs of dots. In each pair, a moving dot collided at speed s into another dot at a standstill, which then started to move at speed s'. They showed that subjects were quicker to spot pairs in which the ratio s/s' was 3/1 than pairs in which it was 1/3, and suggested that the reason has to do with Newtonian mechanics: a ratio of 3/1 is consistent with causal laws of elastic collision, whereas ratios of 1/3 are not. In this case, subjects seem to take the dots to be indicative of events that obey certain causal rules of the external world.

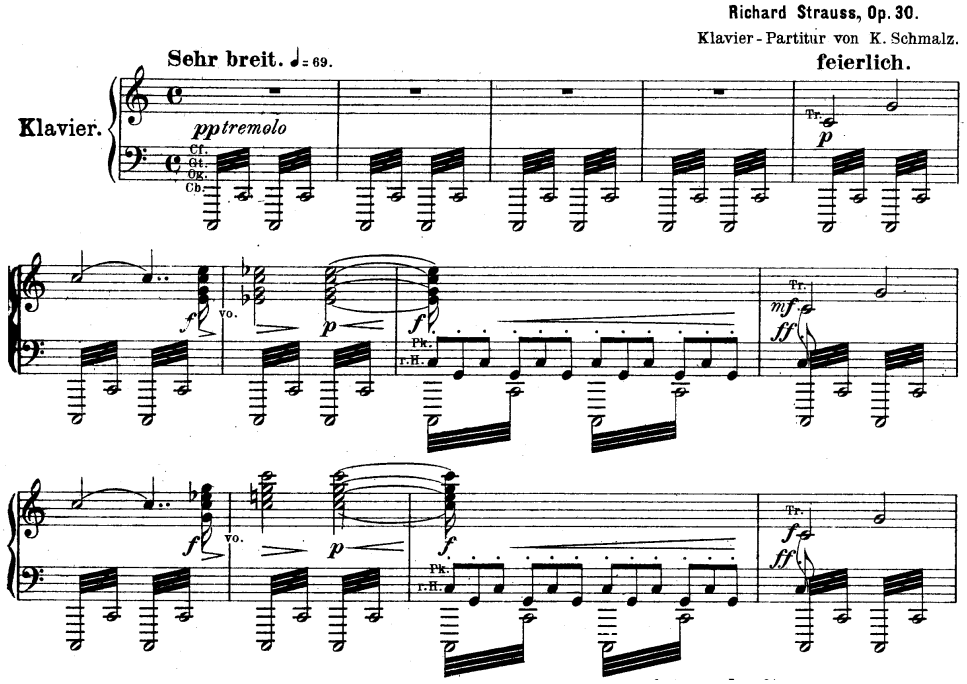
In the visual domain, then, very abstract shapes can still give rise to inferences about virtual events that they are the 'visual traces' of.

3.2 Musical example

Let us turn to music, where sounds will play the role of 'auditory traces' of virtual events. Since the Null Theory is so plausible, we will start by giving an example in which semantic inferences are drawn as well. While they are quite abstract, we believe that they are genuinely semantic, in the sense that pertain to the development of phenomena in the extra-musical world.

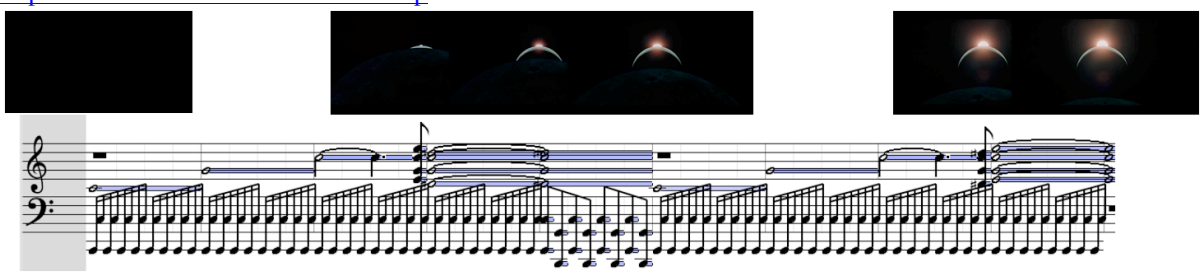
We start with the beginning of Strauss's Zarathustra ('Sunrise'), which is used as the sound track of the beginning of the movie *2001: a Space Odyssey* [https://www.youtube.com/watch?v=e-QFj59PON4&t=0,14s]. A piano reduction is shown in (14).

(14) Beginning of Strauss's *Zarthustra* – piano reduction by K. Schmalz



In (15), we have superimposed some of the key images of the movie to an even simpler reduction (by WilliamWallace); the correspondence already gives a hint as to the inferences one can draw from the music.

(15) Beginning of Strauss's Zarathutra, with the visuals of *2001: a Space Odyssey*
http://www.8notes.com/scores/7213.asp



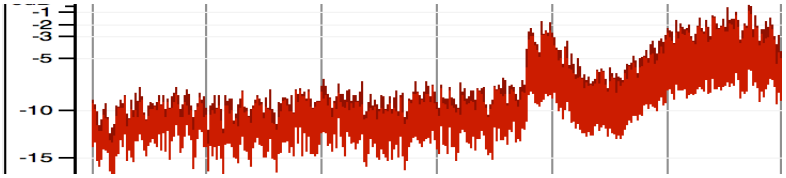
Specifically, the film synchronizes with the music the appearance of a sun behind a planet, in two stages. Bars 1-5 correspond to the appearance of the first half of the sun, bars 5-8 to the appearance of the second half. Now the music certainly evokes the development of a phenomenon in two stages – which is unsurprising as it is an antecedent-consequent structure. But there are interesting twists. A hearer might get the impression that there is gradual development and a marked retreat at the end of the first part, followed by a more assertive development in the second part, reaching its (first) climax in bar 5. Several factors conspire to produce this impression. Three are mentioned in (16). In (16)a, we use chord notation to represent the harmonic development. In (16)b we use numbers from 1 through 5 to represent the melodic movement among 5 different levels (with 1 = lower C, 2 = G, 3 = higher C, 4 = Eb, 5 = E).[4] Finally, in (16)a we use standard dynamics notation to encode loudness (as it can be seen in Schmalz's notation and in the film soundtrack mentioned above).

[4] We set aside the high G and A that appear in measures 10 and 11 in Schmalz's reduction.

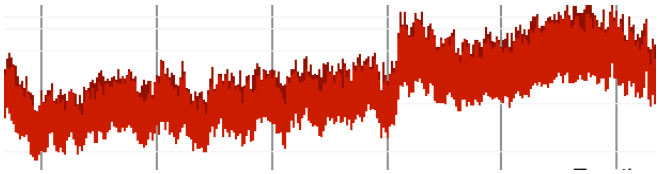
(16) a. Harmony: I – V – I – I Major chord – I minor chord
I – V – I – I minor chord – I Major chord

b. Melody (soparno) 1 2 3 5 4
1 2 3 5 5

c. Loudness: p f > p < f



mf f > p < f



• Harmonically, both the antecedent and the consequent display a movement from degrees I to V to I, but the antecedent ends with a I Major – I minor sequence, whereas the consequent ends with a I minor – I Major sequence. The I minor chord is usually considered less stable than the I Major chord. This produces the impression that there is a retreat at the end of the antecedent, as it reaches a stable position (I Major) and immediately goes to a less stable position (I minor); the end of the consequent displays the opposite movement, reaching the more stable position.

• Melodically, the soprano voice gradually goes up in the antecedent, but then goes down by a half-step at the very end – hence also an impression of retreat. Here too, the opposite movement is found at the end of the consequent.

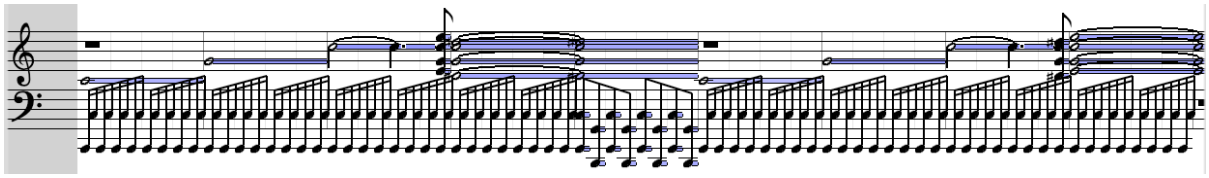
• In terms of loudness, the antecedent starts piano (p), whereas the consequent starts mezzo forte (mf), hence the impression that the consequent is more assertive than the antecedent. Each movement proceeds crescendo, which produces the impression of a gradual development. Finally, each movement ends with a quick decrescendo followed by a strong crescendo, which may give the impression of a goal-directed development, with sharp boundaries in each case.

There would definitely be more subtle effects to discuss.[5] But even at this point, it is worth asking whether harmonic and melodic movement are *both* crucial to the observed impression, in particular that the development retreats at the end of the antecedent. The question can be addressed by determining whether the effect remains when (i) the harmony is kept constant but the melodic movement of the soprano is removed, and (ii) the melodic movement is retained but the harmony is removed.

The simplest way to test (i) is to remove notes responsible for the upward or downward melodic movement while keeping the harmony constant. This is done on the basis of the very simple piano reduction in (17), further simplified to (18)a. In (18)b, two (highlighted) E's responsible for the melodic movement were removed. The initial effect (unstable ending at the end of the antecedent, stable ending at the end of the consequent) is still largely preserved. This might in part be due to the fact that the harmonics of the remaining E's produce the illusion of the same melodic movement as before. But the semantic effect observed is arguably weakened when these remaining E's are lowered by one octave, as is see in (18)c. Clearly, harmony plays an important role in the semantic effect we observe, but the melodic movement seems to play a role as well.

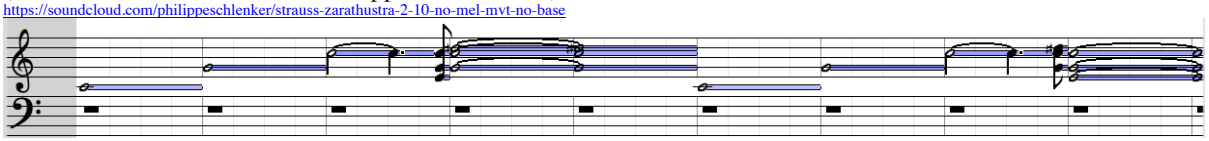
[5] To mention just a few:
• The bars in (15) are not the actual beginning of the piece, but are preceded by 4 bars of tremolo on I (C).
• In both the antecedent and the consequent, the first bar and most of the second bar is underspecified between Major and minor.
These properties should of course be taken into account in a more fine-grained semantic analysis.

(17) A 'bare bones' piano reduction Strauss's Zarathustra, measures 5-13
https://soundcloud.com/philippeschlenker/strauss-zarathustra-2-10-standard?in=philippeschlenker/sets/prolegomena-to-music-semantics

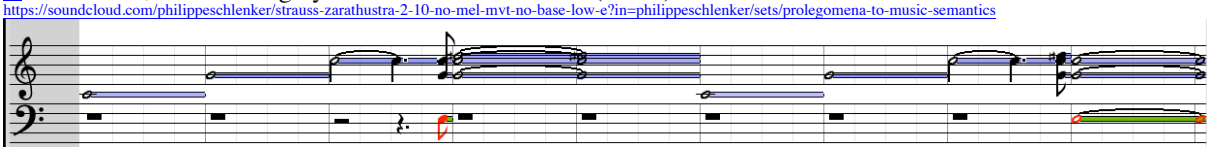


(18) a. Same as (17), without lower voice
https://soundcloud.com/philippeschlenker/strauss-zarathustra-2-10-standard-no-base?in=philippeschlenker/sets/prolegomena-to-music-semantics

b. Same as a., but removing notes responsible for the downward or upward movement of the soprano in a. (the notes that were removed are appear in red in a.)
https://soundcloud.com/philippeschlenker/strauss-zarathustra-2-10-no-mel-mvt-no-base

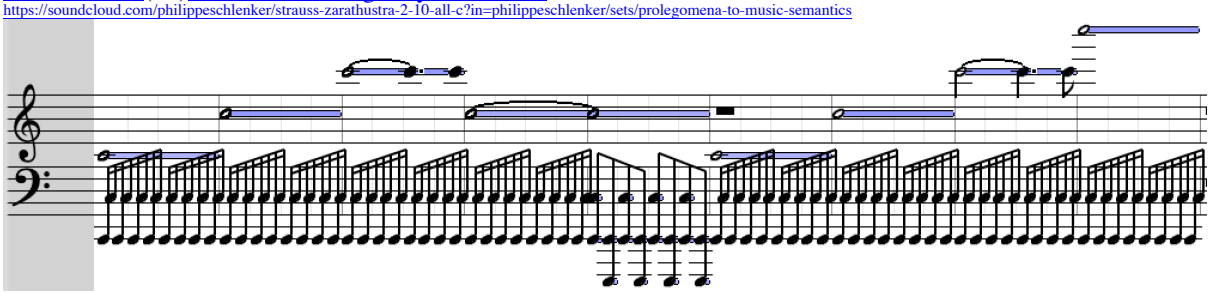


c. Same as b., but lowering by one octave the lower Es (in red)
https://soundcloud.com/philippeschlenker/strauss-zarathustra-2-10-no-mel-mvt-no-base-low-e?in=philippeschlenker/sets/prolegomena-to-music-semantics



The contribution of the melodic movement can be further highlighted by turning to (ii) and asking what effect is obtained if we rewrite (18)a so that only the note C is used, going one octave up or one octave down depending on the melodic movement. What is striking about the result is that it strongly preserves the impression of a two-stage development, with a retreat at the end of a first stage and a more successful development at the end.

(19) A version of (18)a re-written using only the note C
https://soundcloud.com/philippeschlenker/strauss-zarathustra-2-10-all-c?in=philippeschlenker/sets/prolegomena-to-music-semantics



Thus in this case harmonic and non-harmonic properties conspire to yield a powerful effect, and their respective contributions can be isolated by rewriting the piece in various ways. Why should one draw inferences on the basis of loudness and (non-harmonic) pitch height? As a first approximation, we can note that in normal auditory cognition a source of noise can usually be inferred to have more energy if it is louder; and given a fixed source, if the frequency increases, so does the number of cycles per time unit, and hence also the level of energy. On the tonal side, normal auditory cognition won't be directly helpful to draw inferences, and it seems that in this case stability properties of tonal pitch space are put in correspondence with stability properties of real world events. The challenge will thus be twofold. First, we should establish more systematically that inferences are indeed drawn on the basis of normal auditory cognition on the one hand, and of properties of movement in tonal pitch space on the other; we will attempt to do so in Sections 4 and 5. Second, we should develop a framework in which both types of inferences can somehow be aggregated; we will sketch one in Section 6.

4 Semantic effects I: inferences from normal auditory cognition

Noise gives rise to all sorts of inferences about the sources that caused it.[6] In this section, we focus on inferences about virtual sources of the music that one can draw on the basis of normal (non-musical) auditory cognition. We will assume that the sources have been identified (by voice leading principles of classical music theory, or by principles of Auditory Scene Analysis applied to music), and as a first approximation we will take the inferences to pertain to the virtual sources of these voices. (In a more sophisticated analysis, one could explore more subtle musical mechanisms that produce the impression of a background or even of an atmosphere.[7] We briefly come back to related issues in Section 9 when we discuss the role of emotions in music semantics.)

4.1 Timber

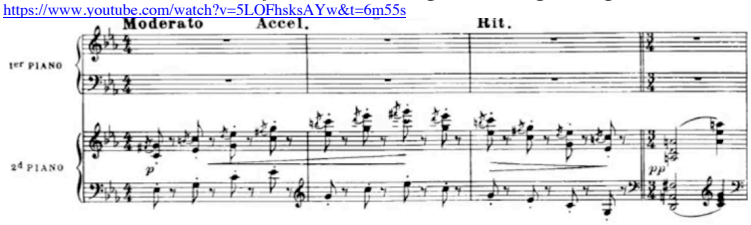
While this might be too obvious to be noted, timber can give an indication about the identification of the voices. This is especially the case when different timbers can be clearly separated in the auditory stream. Systematic use of this device is for instance made in Prokofiev's *Peter and the Wolf,* where the wolf is represented by the sound of French horns, Peter by the strings, the bird by the flute, the grandfather by the bassoon, etc. Even in this case, the mapping isn't trivial, since all the strings together represent Peter. In many other cases, the mapping could be quite a complex, and timber wouldn't necessarily help: a single piano typically produces notes that correspond to several voices; and in orchestral music a single voice is often realized by a combination of timbers.[8]

4.2 Sound and silence

Continuing with the obvious, sound is taken to reflect the fact that something is happening to the source, while absence of sound is interpreted as an interruption of activity or the disappearance of the source. Which entails that the number of sound events per time unit will give an indication of the rate of activity of the source.

A very simple illustration of this effect can be found in Saint-Saëns's *Carnival of the Animals* (Saint-Saëns 1886), in the part devoted to kangaroos [https://www.youtube.com/watch?v=5LOFhsksAYw&t=6m55s]. When the first piano enters, it plays a series of fifth notes separated by fifth silences. This evokes a succession of brief events separated by interruptions. In the context of Saint-Saëns's piece, one can interpret these sequences as evoking kangaroo jumps: for each jump, the ground is hit, hence a brief note, and then the kangaroo rebounds, hence a brief silence. The inferences obtained would be more abstract if we didn't have the title and context of the piece, of course, but the main effect would remain, that of a succession of brief, interrupted events.

(20) Saint-Saëns's *Carnival of Animals,* Kangaroos, beginning
https://www.youtube.com/watch?v=5LOFhsksAYw&t=6m55s



4.3 Speed and speed modifications

Since sound (as opposed to silence) provides information about events undergone by the source, changes in the speed of appearance of sound will be interpreted as changes in the rate of appearance of the relevant events. In the quoted piece on kangaroos (in (20)), each series of jumps starts slow, then, accelerates, and then ends slow – and this produces the impression of corresponding changes of

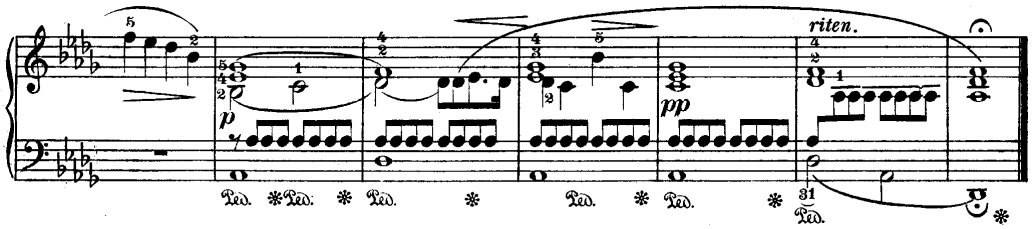
[6] A sound effect evoking a car crash can be found here [https://www.youtube.com/watch?v=PTfKEyqTjzI]. It is striking that the sound alone gives a precise indication of the nature and development of the event.
[7] A similar distinction is needed for simple noise: we may perceive a car as approaching within a background of road-related noises that might not be as distinct. Similarly, an animal's call may be perceived in a background of other noises, for instance of the rain falling or of the wind blowing.
[8] Intrinsic properties of a given timber (rather than just its distinction from other timbers) can certainly trigger inferences as well. This is clearly the case in iconic music in which a sound is to resemble the thing it evokes, as in Saint-Saëns's use of the clarinet to represent cuckoos [https://www.youtube.com/watch?v=5LOFhsksAYw&t=10m35s] or of flutes to represent an aviary [https://www.youtube.com/watch?v=5LOFhsksAYw&t=12m40s]. But some semantic effects are more subtle – as in Saint-Saëns's use of flutes in the melody intended to evoke an aquarium [https://www.youtube.com/watch?v=5LOFhsksAYw&t=7m51s]. Presumably the smooth and continuous sound produced by the flute helps evoke the movement of a marine animal; the less continuous sound of a piano would be less apt to do so.

speed in the kangaroos' jumps (see for instance Eitan and Granot 2006 for experimental data on the connection between 'inter-onset interval' and the scenes evoked in listeners).

The tempo of an entire piece can itself have semantic implications. An amusing example can be heard in Saint-Saëns's tortoises [https://www.youtube.com/watch?v=5LOFhsksAYw&t=3m21s]. It features an extremely slow version of a famous dance made popular in an opera by Offenbach (the 'infernal galop' [https://www.youtube.com/watch?v=okQRnHvw3is&t=1m48s]). Saint-Saëns's version evokes very slow moving objects that attempt a famous dance at their own, non-standard, pace. Similarly, Mahler's Frère Jacques [https://soundcloud.com/philippeschlenker/mahler-frere-jacques-1-6-normal/s-cyGju] departs from the 'standard' Frère Jacques not just in being in minor key (and in some melodic respects), but also in being very slow – which is important to evoke a funeral procession. A version of a midi file in which the speed has been multiplied by 2.5 [https://soundcloud.com/philippeschlenker/mahler-frere-jacques-1-6-speed25/s-OG4VF] loses much of the solemnity of Mahler's version – and it also sounds significantly happier – a point to which we return in Section 9.2.1.

There are certainly more abstract effects associated with speed. In our experience of the non-musical world, speed acceleration is associated with increases in energy, and conversely deceleration is associated with energy loss. This is probably the reason why it is customary to signal the end of certain pieces with a deceleration or 'final ritard' (see Desain and Honing 1996). An example among many others if Chopin's 'Raindrop' Prelude, which features an 'ostinato' repetition of simple notes – which could be likened to raindrops hitting a surface. The last two bars include a strong ritenuto. Artificially removing it weakens the impression that a natural phenomenon is gradually dying out (for reasons we will come to shortly, there are several other mechanisms that also yield the same impression, hence just removing the speed change doesn't entirely remove the impression but just weakens it).

(21) Last bars of Chopin's Prelude 15 ('Raindrop')
a. The last two bars include a ritenuto (normal version).
https://soundcloud.com/philippeschlenker/chopin-prelude-15-last-2-bars-normal



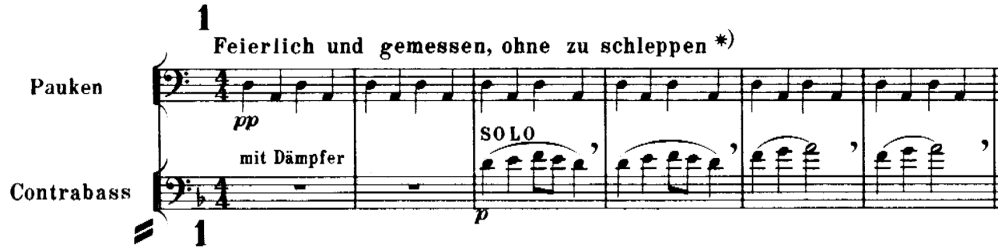
b. A modified version of a. with constant speed in the last two bars does not yield the same impression of a phenomenon gradually dying out.
https://soundcloud.com/philippeschlenker/chopin-prelude-15-last-2-bars-no-rit

In addition, sources that are analyzed as being animate can be thought to observe an 'urgency code' by which greater threats are associated with faster production rates of alarm calls (e.g. Lemasson et al. 2010). This presumably accounts for the association of greater speeds with greater arousal (we come back to related issues in Section 9.2.1).

4.4 Loudness

A noise that is becoming louder could typically be interpreted in one of two ways: either the source is producing the noise with greater energy, or the source is approaching the hearer (see Eitan and Granot 2006). The first case is pervasive in music. The second case can be illustrated by manipulating the loudness of a well-known example. The beginning of Mahler's (minor version of) Frère Jacques (First Symphony, 3rd movement) starts with the timpani giving the beat, and then the contrabass playing the melody, all pianissimo, as shown in (22)a. One can artificially add a marked crescendo to the entire development – and one plausible interpretation becomes that of a procession (possibly playing funeral music, as intended by Mahler) which is gradually approaching.

(22) Mahler's Frère Jacques (First Symphony, 3rd movement)
a. Beginning, normal version
https://soundcloud.com/philippeschlenker/mahler-frere-jacques-1-6-normal-1



b. Beginning, with an artificially added crescendo: this can yield the impression that a procession is approaching
https://soundcloud.com/philippeschlenker/mahler-frere-jacques-1-6-crescendo-beg

c. End: depending on the realization, the decrescendo might be indicative of a procession moving away.
https://soundcloud.com/philippeschlenker/mahler-frere-jacques-last-6-normal

Without any manipulation, the end of Mahler's Frère Jacques displays a decrescendo that can probably be interpreted as the source gradually losing energy, but which can also plausibly be interpreted as a procession moving away from the perspectival center.[9]

Interestingly, just looking at the interaction between speed and loudness, we can begin to predict how an ending will be interpreted. As noted, a diminuendo ending can be interpreted as involving a source moving away, or as a source losing energy. In the former case, one would not expect the perceived speed of events to be significantly affected.[10] In the second case, by contrast, both the loudness and speed should be affected. The effect can be tested by exaggerating the diminuendo at the end of Chopin's Raindrop Prelude in (21); without the ritenuto, the source is easily perceived as moving away.[11]

(23) Last bars of Chopin's Prelude 15 ('Raindrop')
a. Exaggerated version of the diminuendo in the normal version, with a ritenuto
https://soundcloud.com/philippeschlenker/chopin-prelude-15-last-2-bars-normal-dim
The source seems to gradually lose energy, becoming slower and softer.
b. Same as a., but without ritenuto
https://soundcloud.com/philippeschlenker/chopin-prelude-15-last-2-bars-normal-dim
The source seems to be moving away, as it gradually becomes softer, without change of speed.

4.5 Pitch Height

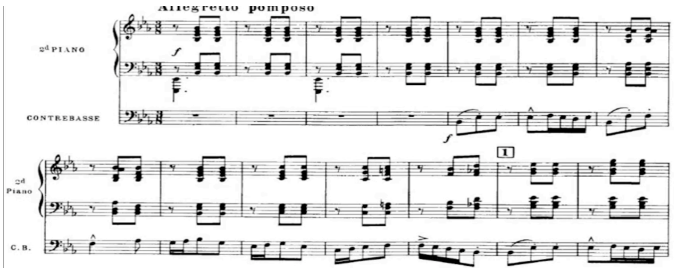
Pitch plays a crucial role in the tonal aspects of music. But keeping the melody and harmony constant, pitch can have powerful effects as well, which we take to be due to the inferences that it licenses about the (virtual) source of the sound. Two kinds of inferences are particularly salient.

(i) The register of a given source – especially for animals – provides information about its size: larger sources tend to produce sounds with lower frequencies.[12] This is a sufficiently important inference that some animals apparently evolved mechanisms – specifically, laryngeal descent – to lower their vocal-tract resonant frequencies so as to exaggerate their perceived size (Fitch and Reby 2001). This is put to comical effect in Saint-Saëns's Carnival, where the melody of a dance is played with a double bass to figure an elephant [https://www.youtube.com/watch?v=5LOFhsksAYw&t=5m24s]. The specific effect of pitch, keeping everything else constant, can be seen by comparing Saint Saëns's version (in a midi rendition, as in (24)a) to an artificially altered version in which the double bass part was raised by two octaves. The

[9] Effects of distance can be produced by literally placing the instruments further away from the listener, as is for instance done in Mahler's Second Symphony, e.g. with 'horns in the distance' [https://www.youtube.com/watch?v=DUD3mrekxtk].
[10] Note that in normal auditory cognition the perceived event rate can marginally be affected when a source moves away. In particular, the apparent frequency of a sound moves down when a source moves away (Doppler effect). It is not clear that this phenomenon plays a role in music. In fact, Eitan and Granot 2006 note that pitch rises are 'significantly associated with moving away', which is the opposite of what the Doppler effect would lead one to expect, as they note.
[11] If we add a crude crescendo instead, and a final accent, the ending sounds more intentional, as if the source were gradually gaining stamina as it approaches its goal, and marks the latter with a triumphant spike of energy [https://soundcloud.com/philippeschlenker/chopin-prelude-15-last-2-bars-cresaccent]. An intentional, triumphant effect is often produced in fortissimo endings, e.g. at the end of Beethoven's Symphony 8 [https://www.youtube.com/watch?v=C2Avpt9FKP0&t=26m10s].
[12] As Cross and Woodruff 2008 note, this correlation lies at the source of a 'frequency' code', discussed in linguistics by Ohala 1994, according to which lower pitch is associated with larger body size.

impression that a large animal is evoked immediately disappears. If the double bass part if raised by 3 octaves, we even get, if anything the evocation of a small source (as in (24)c).[13]

(24) Saint-Saëns's Carnival of Animals, The Elephant, beginning
a. The normal version features a double bass to evoke a large animal.
https://soundcloud.com/philippeschlenker/saint-saens-carnival-elephant-normal

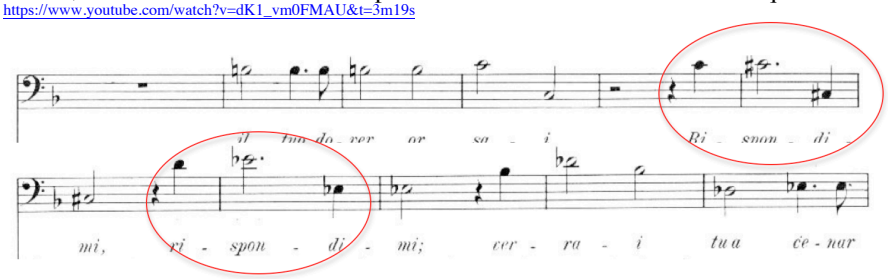


b. Raising the double bass part by 2 octaves (while leaving the piano accompaniment unchanged) removes the evocation of a large source.
https://soundcloud.com/philippeschlenker/saint-saens-carnival-elephant-2-oct

c. Raising the double bass part by 3 octaves might even evoke a small rather than a large source.
https://soundcloud.com/philippeschlenker/saint-saens-carnival-elephant-3-oct

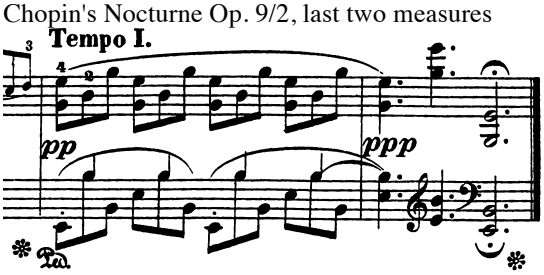
(ii) For a given source, higher pitch is associated with a source that produces more events per time units, hence might have more energy or be more excited. We already saw a version of this effect in the version rewritten only with C notes of the beginning of Strauss's *Zarathustra* in (19). A chromatic ascension with repetition is also used in the Commendatore scene of Mozart's Don Giovanni to highlight the increasingly pressing nature of the Commendatore's order: *rispondimi! rispondimi!* ('answer me! answer me!'; it probably tends to be produced crescendo, which of course adds to the effect).

(25) Mozart's Don Giovanni, Commendatore scene, 'Rispondimi': repetition is produced with a chromatic ascent, which contributes to the impression that the Commendatore's request is becoming more pressing
https://www.youtube.com/watch?v=dK1_vm0FMAU&t=3m19s



If these remarks are on the right track, all other things being equal, the end of a piece should sound slightly more conclusive if the last movement is downward rather than upward. This effect can be found at the end of Chopin's Nocturne Op. 9/2, which ends with two identical chords, except that the second is 2 octaves below the first one. If the score is re-written so that the piece ends upwards rather than downwards, the effect is a bit less conclusive, as is illustrated in (26) (see Appendix I for a similar effect and manipulation at the end of Beethoven's Third Symphony).

(26) Chopin's Nocturne Op. 9/2, last two measures



a. The original version ends with two identical chords, the second one 2 octaves below the first one.

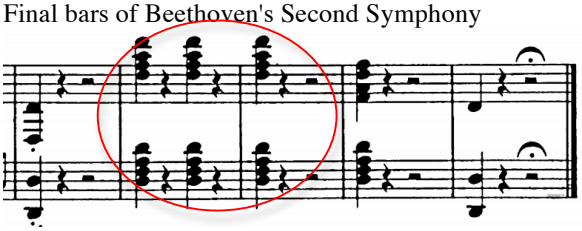
[13] The same property is used throughout opera, where for obvious reasons bass registers are associated with males, but can also be used to indicate power, as in the appearance of the statue of the Commendatore in Mozart's *Don Giovanni*.

https://soundcloud.com/philippeschlenker/chopin-op9-2-better-115-end

b. If instead the second chord is raised by 3 octaves and thus ends up being 1 octave above the first one, the effect is less conclusive.
https://soundcloud.com/philippeschlenker/chopin-op9-2-better-115-end-2-octaves-up

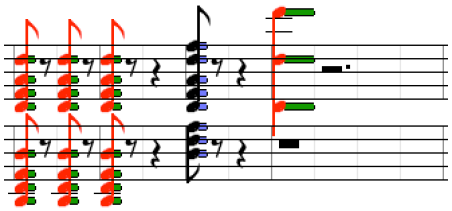
In (26), raising the last chord had the effect of simultaneously changing the direction of movement of the melody: in (26), the last three chords are respectively medium, high and low in the original version, and movement that gets turned into medium, high and higher after the last chord is raised. The end of Beethoven's Second Symphony provides an example in which the final movement ends entirely downwards, as shown (in Liszt's piano transcription) in (27)[14]: the last three chords are respectively at levels high, medium and low.

(27) Final bars of Beethoven's Second Symphony

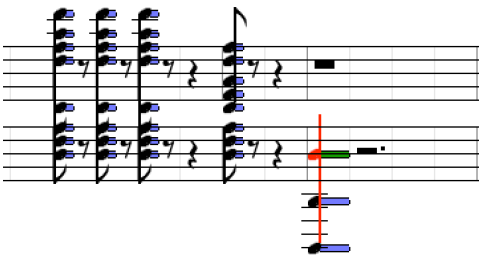


We can see the effect of the direction of movement by lowering the circled chords in (27) by one octave, and raising the final chord by two octaves, as shown in (28)b; the effect is far less conclusive than the original in (28)a. Conversely, we can also increase the effect found in the original piece by lowering the last chord even more. Despite the fact that this creates a greater discontinuity between the penultimate and the last chord, the effect is at least as conclusive as in the original version.

(28) Final bars of Beethoven's Second Symphony
a. The original version sounds conclusive, with a descending melodic movement
https://soundcloud.com/philippeschlenker/beethoven-2nd-symphony-liszt-440-end-normal
b. Lowering the circled chords in (27) by one octave and raising the final chord by two, yields a rising movement which is less conclusive
https://soundcloud.com/philippeschlenker/beethoven-2nd-symphony-liszt-440-end-upwards



c. Lowering the final chord by one octave in (27)a yields a conclusive movement, possibly more so than the original version.
https://soundcloud.com/philippeschlenker/beethoven-2nd-symphony-liszt-440-end-downwards-lower



4.6 Iconicity

As should be obvious, some inferences about the sources of the music are drawn because the music resembles certain sounds we know from our normal auditory experience; these are thus 'iconic' effects. Saint-Saëns's Carnival has a clarinet off-stage evoking a cuckoo by way of a series of descending two-note sequences in The Cuckoo in the Depths of the Woods [https://www.youtube.com/watch?v=5LOFhsksAYw&t=10m35s]. Here timber, frequency and spatial origin of the sound conspire to yield a strong evocative effect. Tchaikovsky's 1812 Overture makes heavy use of iconic means as well, simultaneously using the Marseillaise and the sound of cannons (written into the score) to represent retreating French armies [https://www.youtube.com/watch?v=ZrsYD46W1U0&t=12m39s]. Famously, the Star-spangled

[14] Thanks to Arthur Bonetto for suggesting that we consider this example.

Banner is a recurring theme of Puccini's Madam Butterfly [https://www.youtube.com/watch?v=YgLQ3p6hSOI], where it serves to evoke the American navy (it is only in later years that it became the US national anthem). Finally, piano students doing scales [https://www.youtube.com/watch?v=5LOFhsksAYw&t=13m51s] – with abominable errors – belong to the menagerie described in Saint Saëns's *Carnival*.

The effects we described in earlier paragraphs are arguably quite general; the iconic effects mentioned here are not, and are thus of lesser interest. Still, it would be desirable for a music semantics to derive these rather special cases without stipulations. The source-based analysis straightforwardly delivers this result: these are simply cases in which inferences are drawn as if the sounds were not in a musical context. Thus the sound of a cannon is attributed to a virtual source which is a cannon, and a scale with errors is attributed to a piano student's hapless practice.

4.7 Extensions

In this piece, we only scratch the surface of the mechanisms that trigger inferences on the basis of normal auditory cognition. Some were investigated with experimental means by Eitan and Granot 2006, who asked subjects to imagine the motion of cartoon character that corresponded with various melodic figures. They found that "aspects of timing (IOI and motivic pace) were strongly related to speed (...); aspects of pitch contour were related to verticality, in line with the Western tradition of notation and musical discourse (...); and aspects of loudness were related to distance and energy". Their conclusions on timing and loudness correspond exactly to those that were discussed above. The connection between pitch and verticality is less easy to analyze from the present perspective, and would require a more detailed study.

Rather than delving more deeply into a topic we must leave for future research, we will give one example which involves several factors at once. Consider repetitions. Performers know that any repeated motive leads to crucial decisions concerning its execution. In fact, we already saw several relevant examples.

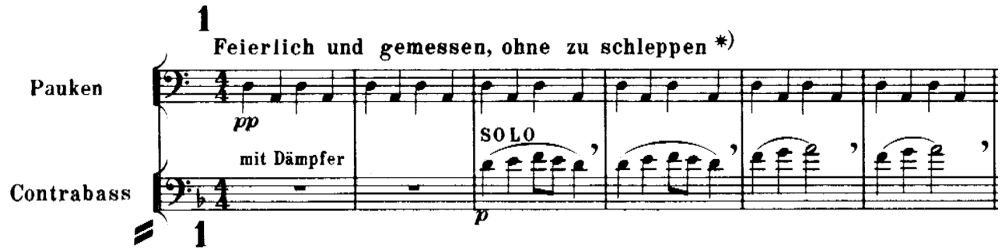
• The last notes of Mahler's Frère Jacques involve a repetition with attenuation of the loudness, and in a standard version [https://soundcloud.com/user-985799021-177497631/mahler-frere-jacques-last-6-normal/s-qCUPT] they could be interpreted in terms of a source moving away, or gradually dying out. But if a strong rallentando is added [https://soundcloud.com/user-985799021-177497631/mahler-frere-jacques-last-6-ralent/s-qdHOL], the 'moving away' interpretation becomes less likely, and the 'dying out' interpretation becomes more salient – which is exactly the effect we discussed in connection with the end of Chopin's Raindrop Prelude in (26).

• Outside of endings, repetitions can be interpreted very differently as well depending on how they are realized. The Commendatore's *rispondimi* (in (25) above) is both linguistically and musically interpreted as a reiteration because the second iteration is chromatically higher than the first, and at least as loud.

• We can also manipulate the beginning of Mahler's Frère Jacques to modify the interpretation of the initial repetitions. A repetition which is realized far more softly that its antecedent may sound like an echo of it, as in (29)b. Louder realization of the repetition may be interpreted as re-assertion, or possibly as a dialogue between two voices, as in (29)c.

(29) Mahler's Frère Jacques (First Symphony, 3rd movement)

a. Beginning, normal version
https://soundcloud.com/user-985799021-177497631/mahler-frere-jacques-3-6-normal/s-10XsM



b. If measures 4 and 6 are realized far less loudly than measures 3 and 5, one can obtain the impression of an echo, or of a dialogue between two voices, one of which is in the distance.
https://soundcloud.com/user-985799021-177497631/mahler-frere-jacques-3-6-echo-30/s-CxzyO

c. If measures 4 and 6 are realized far more loudly than measures 3 and 5, one can also obtain the impression of a dialogue between two voices, or one can get the impression that measures 3 and 5 are reasserted more strongly by the same voice.
https://soundcloud.com/user-985799021-177497631/mahler-frere-jacques-3-6-echo30/s-zXeUA

The key is that repetitions are rarely the product of chance. Depending on how they are realized, they may thus yield the inference that a phenomenon is naturally repeating itself, often with loss of energy and thus attenuation – unless the source is approaching the perspectival center, in

which case the perceived level of energy may increase. Alternatively, the source may be intentional and may be reiterating an action something that wasn't initially successful, possibly with more energy than the first time around. Yet another possibility is that one source is imitating another one, hence the impression of a kind of dialogue. The typology will no doubt have to be enriched.

4.8 Methods to test inferences from normal auditory cognition

Our list of inferences drawn from normal auditory cognition is only illustrative, and ought to be expanded in future research. We believe that such inferences could be tested by using the following method (but see Eitan and Granot 2006 for a discussion of methods developed to test more specifically the relation between music and movement).

1. First, a clear hypothesis should be stated – for instance that, all other things being equal, a given source will be inferred to have greater energy when it produces a higher than a lower sound.

2. Second, minimal pairs should be constructed to assess the inference in a musical context. This could be done in two ways.
(i) One may select actual musical examples, and manipulate them so as to get contrasting pairs, as we did with the end of Chopin's Nocturne 9/2 (in (26)) and Beethoven's Second Symphony (in (28)).
(ii) Alternatively, one may create artificial stimuli which also display a minimal contrast with respect to the relevant parameter, but might be simpler than examples from 'real' music, as we did in our discussion of Strauss's Zarathustra (in (19)).

In each case, one should state a target inference about the source, and determine whether it is triggered more strongly by one stimulus or by the other. One may have abstract statements in natural language – e.g. Which of these two pieces sounds more conclusive? or: Which of these two pieces evokes a phenomenon with the greater level of energy? Or one could resort to indirect ways of testing the inference, for instance by having subjects match musical stimuli with non-musical scenes (for instance visual ones). Which types of statements will prove most productive is entirely open as things stand, and it is likely that different methods will have to be developed depending on the particular goals of the research. Finally, semantic intuitions might be sharpened by initially restricting the set of models the subjects consider. This in effect what program music and sometimes just titles do. For instance, one may tell subjects that a piece represents the movement of the sun, and ask them questions about what they infer about that movement at various points in the development of the piece.

3. Third, one will have to show that these inferences are genuinely triggered in non-musical cognition as well.[15] This may be done by creating non-musical stimuli – for instance with noise, or in some cases with human voices or even with animal calls – that make it possible to test the parameter under study. (In some cases one may even go further and suggest that the relevant properties exist across modalities, and have a counterpart in visual cognition).

4. Finally, as we briefly suggested in our discussion of endings and repetitions, a source-based semantics will prove particularly useful when the interaction of several properties is explored, as the inferences will become much richer in this case.

5 Semantic effects II: inferences from tonal properties

5.1 The need for a tonal component

The inferences we discussed so far were all 'lifted' from normal auditory cognition. They could have applied to 'real world' noise, or to auditory versions of Heider and Simmel's abstract cartoons, discussed in Section 3.1: the voices would have been attributed to abstract sources, but the laws that regulate their behavior would have been those that apply to objects and to sound in the normal world. Music is special in that the voices are located in and constrained by a non-standard space, tonal pitch space; they have melodic and harmonic properties that are rarely found in the natural world outside of music-related phenomena. Tonal pitch space comes in different varieties different musical tradition, and even within one and the same musical tradition, as shown by the distinction between major and minor keys. The formal properties of tonal pitch space have been studied in great detail, in particular in the Western classical tradition – and they play a prominent role in studies of musical syntax (see

[15] In fact, one should in the end distinguish between two issues: (i) is an auditory inference physically licensed – given the physics of sound? (ii) is it cognitively licensed? There might be discrepancies between (i) and (ii).

Lerdahl and Jackendoff 1983, Lerdahl 2001, Granroth-Wilding and Steedman 2014). Here we will be interested in the role that tonal properties should play in a music semantics.

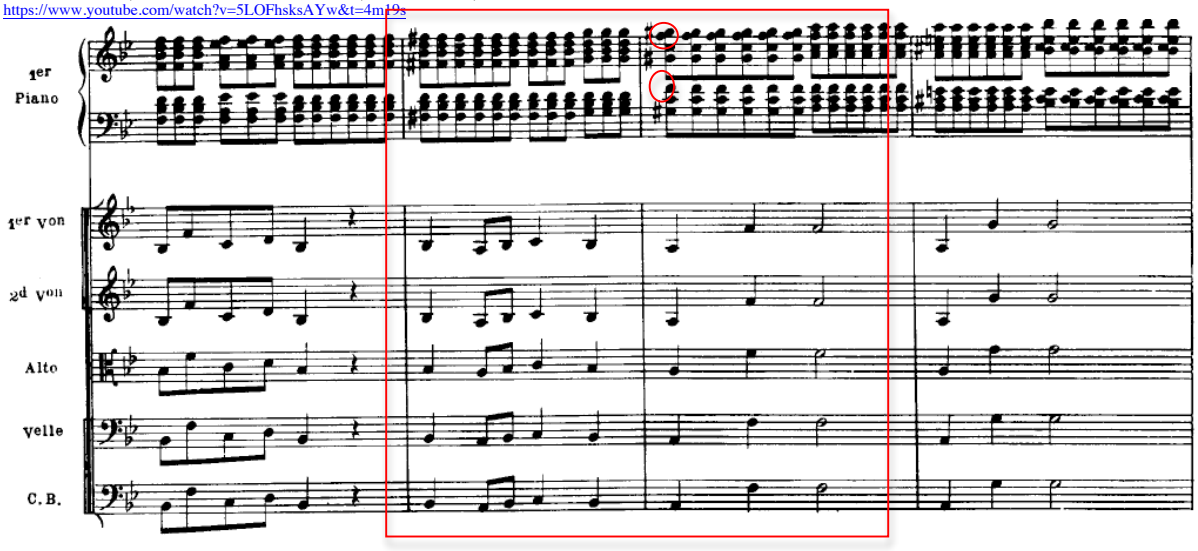
As will be recalled, we introduced at the outset the metaphor of objects seen on a roller-coaster in the distance, with the idea that one needs to take into account the special physical properties of the roller-coaster to draw appropriate inferences about the objects that interact with it. Similarly, in order to draw inferences about the sources of the music, one needs to understand how they interact with tonal pitch space: their position may be stable or unstable, they may be attracted to other positions, etc. Of course the limitation of our metaphor is that an understanding of the physical properties of a roller-coaster entire relies on normal visual perception, whereas tonal pitch space largely departs from normal auditory cognition.

A key challenge – to be addressed in Section 6 – will be to aggregate inferences that drawn thanks to normal auditory cognition, and further inferences drawn on the basis of the behavior of musical voices in tonal pitch space. Restricting attention to the inferences drawn from normal auditory cognition (as could be done by applying the considerations of Section 4, or Bregman's auditory scene analysis) would miss what crucially distinguishes music from noise. Conversely, one might be tempted to restrict attention to tonal properties, for instance by taking the meaning of a music to be a journey through tonal pitch space, as is informally suggested by Lerdahl 2001 and formally implemented in in Granroth-Wilding and Steedman 2014. But this would miss the rich inferences which are drawn from normal auditory cognition, several examples of which were discussed in Section 4.

5.2 An example: a dissonance

A very simple example will help illustrate the inferential power of tonal inferences. In Saint Saëns's impossibly slow version of the *Can Can* dance, which he uses to represent tortoises, there are moments of severe dissonance, and they produce a very powerful effect. The very slow dance evokes the tortoise's very slow walk. But when we hear a dissonance in measure 12, copied in (30), we get the impression that the tortoises are tripping on something. In the words of the Calgary Philharmonic Education Series, the dissonances "evoke the scene of lumbering turtles trying to dance and haplessly tripping over their feet."[16] While at first it may seem that the musicians are impossibly out of tune, in fact they are just playing a dissonant chord, with both A and Ab in the same chord, as shown in (30). When the Ab is replaced with A throughout this half-measure (as in (30)b), the dissonance disappears, as does the impression that the tortoises are tripping.

(30) Saint Saëns, Carnival of Animals, Tortoises, measures 10-13
https://www.youtube.com/watch?v=5LOFhsksAYw&t=4m19s



a. In the original version, there is a dissonance in the first half of measure 12 because a chord F A C is played with an Ab added (as can be heard by focusing only on the violin and piano parts, for instance).
https://soundcloud.com/user-985799021-177497631/saint-saens-carnival-tortoises-12-13-normal-piano-50/s-J09Qh

b. The dissonance can be removed by turning the Ab's into A's – and the impression that tortoises disappears (as can be heard by focusing only on the violin and piano part, for instance).
https://soundcloud.com/user-985799021-177497631/saint-saens-carnival-tortoises-12-13-corrected-piano-50/s-PXnB0

[16] *Education Concert Curriculum Guide Extreme Music*, Calgary Philharmonic Orchestra's 2005 - 2006 Education Series

In this very simple example, a point of great *tonal* instability is interpreted as corresponding to an event of great *physical* instability for the tortoises, which correspond to the virtual sources of the voices. In the general case, things are far less specific – and in fact if we disregarded Saint Saëns's title, the inferences we would draw wouldn't be specifically about tortoises. But they would probably still involve a source which is slow (due to the comparison with the speed of the standard Can Can), and also in positions of instability at moments that correspond to the dissonances.

5.3 Cadences

5.3.1 Harmonic conclusions

We discussed above (in Sections 4.3, 4.4 and 4.5) the role that speed, loudness and pitch can have in producing the effect that the end of a piece has been reached; one particularly salient case is that in which the source is losing energy and thus produces gradually softer, slower and lower sounds. But a staple of both traditional and contemporary music theory is that tonal considerations play a prominent role as well: a cadence is the standard way of marking the end of a classical piece, typically by way of a dominant chord (V) (often preceded by a preparation in a 'subdominant' region of tonal pitch space), followed by a tonic chord (I). In addition, there are 'half-cadences' that can signal temporary pauses and call for a continuation. These devices are so central that they are 'hard-wired' in Rohrmeier's (2011) syntax of tonal harmony, as is illustrated with his main rules of 'functional expansion' in (31).[17]

(31) Functional Expansion Rules in Rohrmeier 2011

a. *TR* –⟶ *DR t* (4)
b. *DR* –⟶ *SR d* (5)
c. *TR* –⟶ *TR DR* (6)
d. *XR* –⟶ *XR XR* for any *XR* ∈ R (7)
e. *TR* –⟶ *t* (8)
f. *DR* –⟶ *d* (9)
g. *SR* –⟶ *s* (10)

Key:

TR = Tonic Region (= I)	t = tonic
DR = Dominant Region (= V)	d = dominant
SR = Subdominant Region (= IV)	s = subdominant
XR = any region of type X	

Briefly, every piece will be the expansion of the tonic region *TR*. But the rules in (31) guarantee that each such expansion will end in a dominant – hence a half-cadence; or in a dominant-tonic sequence, hence a full cadence. Specifically, any complex tonic region *TR* will have to be expanded by way of the rule (31)a, yielding *DR t*, where t is a tonic; or by (31)c, yielding TR DR; or by (31)d, yielding *TR TR*. The last case leads back to a phrase ending in a sub-phrase *TR*, hence we can restrict attention to the first two cases.

• Starting with the second case *TR DR*, (31)b, d, f guarantee that *TR DR* will end up being expanded as something that ends in … *d* (where *d* is a dominant) because the expansion of *DR* according to (31)b and (31)f will end in *d*.

• The same reasoning can be applied to the first case *(DR t)* to yield the conclusion that the final expansion will end with something of the form *…d t*. The latter case corresponds to a full cadence, the previous one to that of a half-cadence.

The question that is not addressed in this syntactic framework is *why* certain sequences of chords are used to mark a weak or a strong end. We submit that the traditional intuition, framed in terms of relative stability, is exactly right but needs to be stated within a semantic framework. In brief, a full cadence is final because it ends in a position of tonic space that is maximally stable. A half-cadence is less final because it ends in a position that is relatively stable, but less so than a tonic. Furthermore, cadences are often of the form subdominant - dominant - tonic because this provides a gradual path towards tonal repose, which mirrors one of the patterns we saw with speed and loudness, both of which could be decreased gradually to signal the end of a piece. A semantic analysis couldin principle capture these facts as follows: music is special (compared to noise) in that the sources are understood to exist in a space with very special properties, isomorphic to those of tonal pitch space. In particular, different positions in tonal pitch space come with different degrees of stability, and

[17] This is only part of Rohrmeier's system. More complex patterns are allowed because Functional Expansion Rules are enriched with a set of "substitution rules modelling how functional elements may be substituted by parallels (relatives)", and "two modulation rules formalising modulation and change of mode".

relations of attraction to other positions. As a result, a source can be expected to be in a very stable position if it manifests itself by a tonic chord, and in a less stable, but still relatively stable position if it manifests itself by a dominant.

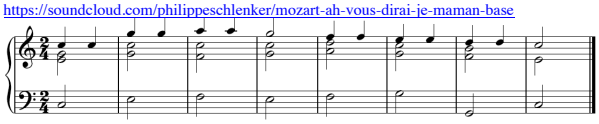
5.3.2 Varieties of cadences

Of course this only scratches the surface of an analysis of cadences Still, the general form of the account seems appropriate to extend to more fine-grained phenomena. To mention just two:
–A cadence is more conclusive if the final tonic chord has a tonic in root position than if it appears in inverted form. This is presumably because in the former case the chord is more stable.
–If the final I chord is replaced with a VI chord (which shares with it 2 out of three notes – e.g. C E G vs. A C E), the result is less stable – hence the term of a 'deceptive cadence'.[18]

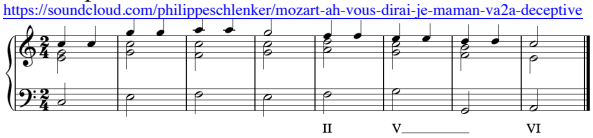
It is worth giving an example of the effect of the slightly 'incomplete' feeling produced by a deceptive cadence. (32)a is a simplified version of the theme of Mozart's *Variations on 'Ah vous dirai-je maman'*. The piece is in C major and the last two measures involve the chords V – I respectively, hence a perfect cadence. In (32)b, only the last two bars are changed, and the melodic line is kept constant, but the harmony is modified so as to obtain a sequence V – VI – hence a 'deceptive' cadence. The effect is considerably less conclusive. By contrast, the same kind of modifications have been made in (32)c, except that now the piece ends in a 'plagal' cadence (here: II – I). The effect seems rather conclusive, possibly as much so as the perfect cadence V – I, and certainly much more so than the deceptive cadence V – VI.

(32) Ah vous dirai-je Maman, simplified from Mozart's theme (b. and c. written by A. Bonetto)

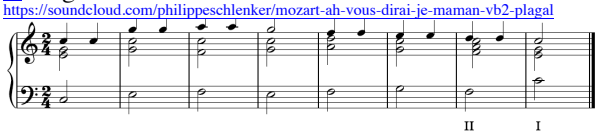
a. Perfect cadence: II V I
https://soundcloud.com/philippeschlenker/mozart-ah-vous-dirai-je-maman-base



b. Deceptive cadence: II V VI
https://soundcloud.com/philippeschlenker/mozart-ah-vous-dirai-je-maman-va2a-deceptive



c. Plagal cadence: II V IV
https://soundcloud.com/philippeschlenker/mozart-ah-vous-dirai-je-maman-vb2-plagal



While the topic of cadences is a staple of traditional and recent approaches to music, we believe that they should be studied within a broader framework in which considerations of harmonic stability are studied in tandem with more or less conclusive effects produced by loudness, speed, melodic line, etc. These various parameters provide different sorts of semantic information: we already saw that loudness and speed modifictions trigger difference inferences, and that they can be combined to yield the effect that a source is gradually dying out or moving away. This typology should be enriched by considering how various types of cadences, which provide information about the stability of the positions are reached, interact with the inferences triggered by loudness and speed, among others.

5.4 Musical meaning cannot be equated with musical tension

In music theory, the notion which is most often given a potentially extra-musical correspondent is that of 'musical tension'. Lerdahl 2001 and Lerdahl and Krumhansl 2007 define and experimentally test a model of musical tension based on four components: (i) "a model of tonal pitch space and all distances within it"; (ii) "a model of hierarchical (prolongational) event structure" (to which we briefly return in Section 7.5); (iii) "a treatment of surface (largely psychoacoustic) dissonance"; and (iv) "a model of melodic (voice-leading) attractions".

We take the model of tonal pitch space that underlies the analysis of musical tension to be crucial to an analysis of music semantics, but *not* because the meaning of music is somehow exhausted by musical tension. Rather, the sources of musical events are understood to be located in a

[18] In Rohrmeier's system, substitution rules are needed to account for deceptive cadences.

space isomorphic to tonal pitch space, and it is for this reason that the relative stability of and attraction relations among these positions are crucial to understand the events undergone by the sources.

To go back to our initial metaphor of objects on a roller-coaster, one could probably ask viewers who observe the scene to assess the degree of tension of one or several objects – or possibly of the entire scene – at various moments. When objects are at a standstill in a point of equilibrium, one would certainly take tension to be at a minimum. When objects are in points of disequilibrium, the tension will certainly be much greater. Furthermore, a detailed analysis of the relative stability of various positions will probably prove important to draw correct inferences about an object that goes through these positions – for instance to determine whether it has an inner force that allows it to go against the path of greatest attraction. But from this it doesn't follow that the informational content of the scene is exhausted by these patterns of tension. We submit that the situation is similar with musical tension: it is useful to an understanding of music semantics, but it does not exhaust it.

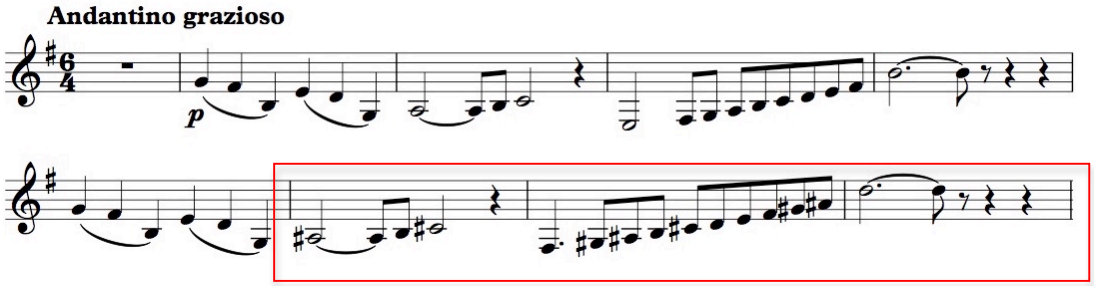
5.5 Modulations

Tonal pitch space is organized into regions, which correspond to keys – with relations of distance among those. Moving to another key that the source is moving towards a new environment (or possibly that one starts perceiving a new source). Furthermore, key change is usually governed by rules of 'modulation', with transitional regions that belong to both keys. This can be seen as a constraint of continuity on possible movements of the sources: a jump to a distant key would be understood as being odd because it would violate this principle.

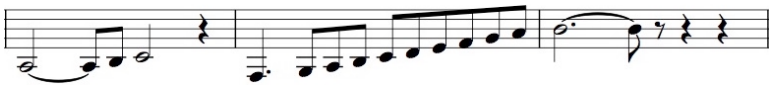
A simple example of a spatial interpretation of a modulation can be found in Saint-Saëns's Swan. The title as well as the initial undulating harp accompaniment are evocative of a movement on water –given the title, that of a swan. The piece is initially G Major but modulates to B minor in measures 7-10, as seen in (33)a. The effect is arguably to suggest the exploration of an area with a different type of landscape. This effect largely disappears if the modulations are rewritten in G Major, as is done in different ways in (33)b-c. Both versions still yield the impression of a movement (on water, given the title and the beginning of the piece), but not so much

(33) Saint-Saëns, The Swan, initial modulation (b. and c. re-written by A. Bonetto)

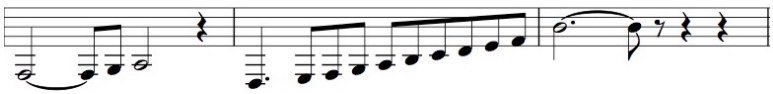
a. Original version, in G major, with a modulation in B minor in measures 7-10
https://soundcloud.com/philippeschlenker/saint-saens-cygne-normal



b. Pure G Major version, with measures 7-9 rewritten by eliminating alterations foreign to G Major, and replacing the final D with a B (to avoid a jump of a fifth between the penultimate and last note)
https://soundcloud.com/philippeschlenker/saint-saens-cygne-v1



c. Pure G Major version, with measure 7-9 rewritten by transposing down (by a third) what is written in B minor (this makes it possible to keep the same melody as in a., one third lower, but in G minor)
https://soundcloud.com/philippeschlenker/saint-saens-cygne-v2



Both rewritten versions preserve the character of a movement, but what gets lost is the impression that a new type of landscape is being explored in measures 7-8.

Another example of a spatial interpretation of modulation is afforded by the soundtrack of Star Trek. It includes abrupt modulations, and these can easily be interpreted as evoking different parts of space (or possibly different adventures in the series) [https://soundcloud.com/philippeschlenker/goldsmith-star-trek-normal]. This impression is largely lost when (a simplified transcription of) the score is rewritten in such a way as to eliminate the modulations [https://soundcloud.com/philippeschlenker/godsmith-star-trek-vb1-14-34]

5.6 Methods and further questions

Having sketched some very simple semantic effects that are triggered by tonal properties of music, we add a word about the methods that could be employed in semantic studies of tonal inferences. In the study of inferences from normal auditory cognition (in Section 4), we could (i) select a semantic effect triggered by a certain property X of the music, and (ii) argue that X gives rise to similar inferences with non-musical stimuli. But because of what tonality is, part (ii) is not applicable. So the analysis must *per force* be more theory-internal. We thus propose that it should include the following steps.

1. First, a hypothesis should be stated – for instance that the leading tone is 'attracted' to the tonic and thus creates an expectation that a voice in the leading tone will then reach the tonic.
2. Second, minimal pairs should be constructed to establish the point. The general methods developed in experimental studies of music could presumably play a role here. In particular, intuitions could be made sharper by restricting the set of models of the music by specifying – by way of a title or a description – what the music is supposed to be about, and then testing semantic inferences that arise given this assumption.
3. Third, instead of correlating these effects with ones that are found in non-musical stimuli, one could seek to explain these effects by properties of tonal pitch space as analyzed (on non-semantic grounds) by the best experimental and formal studies.

Still, although some of the key properties of tonal pitch are not commonly found in normal auditory cognition, one should at some point ask whether ordinary auditory cognition motivates some of the general inferences we draw on the basis of tonal pitch space. We saw for instance that a strong dissonance in tonal pitch space – as in Saint-Saëns's Tortoises – can easily be mapped to an instability in the normal, physical space. What is the basis for this general inference? It would be interesting to explore the inferences that highly dissonant sounds give rise to in *normal* auditory cognition, and possibly use this to motivate the way in which detailed properties of tonal pitch space are semantically interpreted (from this, it does not follow, of course, that one could somehow do without the properties of tonal pitch space in stating a music semantics).

6 Musical truth

We showed in Section 4 that diverse semantic inferences are drawn in music on the basis of properties of normal auditory cognition. We saw in Section 5 that further inferences are drawn on the basis of properties of tonal pitch space. How can these diverse inferences be handled within a unified framework? In principle, this could be done in two ways:

1. **Inferential direction:** we could find a way to simply conjoin all the relevant inferences – and say that *the meaning of a musical piece is the set of inferences it licenses on its sources*.
2. **Model-theoretic direction:** alternatively, we could find a way to explain what it means for a musical piece to be *true* of a situation.

An advantage of the second method is to ensure that the inferences licensed are not contradictory: by providing a situation that makes all of them true, we can be sure that we are not dealing with a system that is trivial because it licenses contradictions. Still, it is often more intuitive to speak of the meaning of music in inferential terms, and it should be emphasized that inferential information will not be lost if we follow the second method. This is because the model-theoretic direction will specify for each musical piece a set of situations (possibly a very large set of very diverse situations) that make it true; the inferences licensed by the music will simply be the properties that are true of all of these situations.

6.1 An example

Because this is all rather abstract, we should start with a highly simplified example. Think again of the C – G – C progression we saw in Strauss's *Zarathustra*, where it was used to evoke a sunrise. We discussed at some length the role played by pitch height, but here we will focus on just two properties, one harmonic and one not:

- within this initial sequence, the key is C (major or minor – this is initially underspecified), and thus C is more stable than G; as a result, the progression is from the most stable position, to a less stable position, back to the most stable position;
- in addition, the progression is realized crescendo.

In order to analyze progressions that just involve these two parameters, we will consider sequences of pairs of the form <note/chord, loudness>, as illustrated in (34). For the sake of generality we take the first members of the pairs to be chords, and we may assume general principles of relative stability of chords, notably the fact that I is more stable than V, which itself is more stable than IV

(within the context of the beginning of Strauss's Zarathustra, one may think instead of different components of a I chord, with C more stable than G).

(34) a. M = <<I, 70db>, <V, 75db>, <I, 80db>>
b. M' = <<I, 70db> , <IV, 75db>, <V, 80db>>
c. M" = <<IV, 80db>, <V, 75db>, <I, 70db>>

So here M is a crescendo progression from I to V to I. M' follows the same crescendo pattern, but goes from I to IV to V; while M" is diminuendo from IV to V to I. For present purposes, a musical piece is just an ordered series of such pairs. The ones we just considered contained only 3 musical events each, but of course there could be more.

Now we will take each pair of the form <note/chord, loudness> to denote an event in the world. Our musical pieces M, M' and M" will thus each depict a series of 3 events in the world. Wolff 2015 argues that the ontology of music involves 'pure' events, but what we have seen before suggests that this will miss some crucial aspects of musical meaning: inferences are derived by considering virtual sources of the voices, and these sources are often identified with *objects in the world*. Accordingly, we associate:
(i) with any voice M an object O;
(ii) with the series of musical events $m_1, \ldots, m_n$ that make up M, a series of world events $e_1, \ldots, e_n$, with the requirement that each of these events should have O as a participant.

(35) Let M a voice, with M = $<M_1, \ldots, M_n>$. A possible denotation for M is a pair $<O, <e_1, \ldots, e_n>>$ of an object and a series of n events, with the requirement that O be a participant in each of $e_1, \ldots, e_n$.

Our model is pictorial in nature: each series of musical events is taken to denote a series of events. Of course at this point nothing specifies whether the representation can be true of the events it depicts.

Thus the next step is to determine under what conditions the depiction is true. In our analysis, this will be the case when these real world events satisfy certain inferences triggered by the musical voice – inferences from normal auditory cognition, and also tonal inferences. Here we will only give a 'toy example' for an analysis of this kind.

Starting from the pieces in (34) and the specification of possible denotations in (35), we will say that the music piece M = $<M_1, \ldots, M_n>$ is true of the pair of an object and events it undergoes, $<O, <e_1, \ldots, e_n>>$, just in case $<O, <e_1, \ldots, e_n>>$ is a possible denotation for M, and in addition the mapping from $<M_1, \ldots, M_n>$ to $<e_1, \ldots, e_n>$ preserves certain requirements, listed in (36).

(36) Truth of
Let M = $<M_1, \ldots, M_n>$ be a voice, and let $<O, <e_1, \ldots, e_n>>$ be a possible denotation for M. **M is true of $<O, <e_1, \ldots, e_n>>$** only if it obeys the following requirements.

a. Time
The temporal ordering of $<M_1, \ldots, M_n>$ should be preserved, i.e. we should have $e_1 < \ldots < e_n$, where < is ordering in time.

b. Loudness
If M_i is less loud than M_k, then either:
(i) O has less energy in e_i than in e_k; or
(ii) O is further from the perceiver in e_i than in e_k.

c. Harmonic stability
If M_i is less harmonically stable than M_k, then O is in a less stable position in e_i than it is in e_k.

While the temporal condition doesn't need justification, the Loudness and Harmonic stability conditions do.
• The preservation condition on Loudness is disjunctive. The intuition is that in auditory cognition in general, softer sounds are associated either with objects that have less energy, or with objects that are further away, as discussed in Section 4.4.
• The preservation condition on Harmonic stability is purely musical, and captures the intuition that less stable events in musical space should denote less stable events in the world. The simplest example of this phenomenon was discussed in Section 5.2 in connection with Saint-Saëns's Tortoises, where a dissonance was rather clearly interpreted as the tortoises tripping.

We can now illustrate how these preservation conditions will lead to a notion of truth. We consider three objects: the sun, a boat, a car. And we will consider 'bare bones' versions of two sequences of events for each. For the sun, a sunrise and a sunset. For the boat, a movement towards the perspectival center, and a movement away from it. For the car, just a car crash. We will analyze these events in a highly simplified fashion, with each event made of three sub-events. In this way, we will obtain five possible denotations for our piece M = <<I, 70db>, <V, 75db>, <I, 80db>> in (34)a.

(37) a. sun-rise = <sun, <minimal-luminosity, rising-luminosity, maximal-luminosity>>
b. sun-set = <sun, <maximal-luminosity, diminishing-luminosity, minimal-luminosity>>
c. boat-approaching = <boat, <maximal-distance, approach, minimal-distance>>
d. boat-departing = <boat, <maximal-distance, departure, maximal-distance>>
e. car-crash <car, <movement_1, movement_2, crash>>

Since M is comprised of three musical events, and each of the sequences in (37) is of the form <object, <event_1, event_2, event_3>>, each is a possible denotation for M according to (35). It remains to see whether M is true of any of these sequences. As we will argue, it should be true of sun-rise and boat-approaching but not of the other events because only sun-rise and boat-approaching involve sequences of events that preserve the key properties of M: the music goes from stable to less stable to more stable (1-5-1); and loudness increases, which can be interpreted as a rise in (real or perceived) level of energy, as in sun-rise, or as an objet approaching the perspectival center, as in boat-approaching.

Let us see in greater detail how this result can be derived. We rely on intuitive properties of the stability or level of energy of events in the world; in a more systematic analysis, some empirical or formal criterion should of course be given to assess 'stability' and 'level of energy' of real world events on independent grounds.

Let us first note that all the sequences of events given in (37) are intended to obey the time ordering condition stated in (36): in each sequence <object, event_1, event_2, event_3>, the events come in the order event_1 < event_2 < event_3. So for M to be true of one of the sequences in (37), all we need to check is that it satisfies the Loudness and the Harmonic Stability conditions.

• Consider first sun-rise in (37)a. Since M is crescendo, M_1 is less loud than M_2, which is less loud than M_3. The Loudness condition in (36)b mandates that minimal-luminosity should have less energy or be further from the perceiver than rising luminosity; and similar for rising-luminosity relative to maximal-luminosity. Certainly the perceived level of energy fits the bill (in physical terms, the interpretation in terms of rising proximity to the perceiver is astronomically correct, though in psychological terms the 'energy'-based interpretation seems more relevant). This shows that the Loudness condition is satisfied. Turning to the Harmonic Stability condition, it too would seem to be satisfied: the initial and final subevents are relatively static, hence stable, whereas the intermediate event is dynamic, hence less stable. In sum, all conditions are satisfied to say that M is true of sun-rise.

• By contrast, we will now see that the same reasoning leads us to say that M is *not* true of sun-set in (37)b. The Harmonic Stability condition is not the issue: just as with sun-rise, the events that begin and end the process can be taken to be the most static and thus stable. On the other hand, the Loudness condition is not satisfied: when we consider the first and the second event, namely maximal-luminosity and diminishing- luminosity, there is neither an increase in 'energy' level, nor an approach.

• Consider now boat-approaching in (37)c. As was the case for sun-set, both the Loudness condition and the Harmonic stability condition will now be satisfied. While it wouldn't make much sense to say that a boat approaching is gaining energy (if anything, it might slow down as it approaches the coast), the sequence corresponding to boat-approaching = <boat, <maximal-distance, approach, minimal-distance>> satisfies the Loudness condition for M because the events in the sequence are getting gradually closer to the perspectival center The ordering $M_1 < M_2 < M_3$ of the crescendo sequence M is thus preserved when M is mapped to the series of events boat-approaching, since the subevents maximal-distance, approach, minimal-distance are similarly ordered in terms of proximity to the perspectival center. Turning to the Harmonic stability condition, the situation is the same as in sun-rise (or for that matter sun-set): the initial and final subevents are the most static, whereas the intermediate event is more dynamic, and less dynamically stable. This corresponds to the ordering by stability of the I V I sequence in M.

• The boat-departing event in (37)d satisfies the Harmonic stability condition, but not the Loudness condition. Its two initial subevents are maximal-distance followed by departure, and the second doesn't have more energy than the first, nor is it closer than it – hence the crescendo character of M is not properly interpreted.

• Finally, the car-crash event in (37)e might or might not satisfy the Loudness condition, depending on whether we take the sequence <movement_1, movement_2, crash> to correspond to an increase in energy and/or to a movement towards the perceiver. But plausibly the Harmonic stability condition is

not satisfied: one would expect that the musical event corresponding to the crash is the least stable of all three events, whereas here it corresponds to the final I of the piece. Things would be different if the piece finished in a highly dissonant chord, but this is not the case here.

In summary, the piece M introduced above is true of sun-rise and boat-approaching but not of the other events considered here.

6.2 Model-theoretic truth vs. Inferential truth

This toy example was given only to illustrate what we take the main components of a music semantics.

• First, we need to specify certain formal properties of the music that must be preserved by the events that the music is true of. Here we isolated three: temporal ordering; relative relations of loudness; and relative relations of stability.

• Second, we must define the set of real world events that the music is taken to be true or false of.

• Third, we must specify under what conditions a series of musical events is true of some real world events.

The last and crucial step could be seen in two ways. One could proceed in an inferential fashion, and take the set of all entailments that can be stated in terms of loudness relations or harmonic stability relations on the musical side, and reinterpret them in terms of energy/remoteness and event stability. Thus one would observe in the case of M in (34)a that M_1 is more harmonically stable than M_2, with a corresponding requirement that the denotation of M_1 be less stable than that of M_2. In effect, this would boil down to reinterpreting with 'real world' vocabulary some musical relations that involve 'musical' vocabulary involving loudness or harmonic stability. Proceeding in this inferential manner, then, we take the content of a musical piece to be the set of inferences it licenses on its virtual sources, where these inferences are obtained by 'translating' musical relations into real world relations in an appropriate way, as illustrated above (greater loudness => greater proximity / greater energy; greater harmonic stability => greater even stability). However this procedure comes with a drawback: when one requires that a set of propositions should be true together, one is not assured that these are not collectively contradictory. To show this, one needs to find a model that satisfies them all. This drawback is avoided in the model-theoretic analysis we have developed in this piece. Instead of defining the set of entailments that must hold of the purported denotations of the musical events, we directly define the class of sequences of real world events of which the musical piece is true. By inspecting this set, we can directly check that the inferences we wish to preserve are not collectively contradictory: they are just in case the set in question is empty.

As we will see shortly (in Section 6.3.2), our analysis of music semantics has the same general structure as a semantics of pictures: if we seek to determine whether a triangle is a correct representation of a particular scene, we seek to map the sides of the triangle to aspects of the scene, and ask whether the mapping preserves key geometric properties of the triangle. This is what we did in a dynamic way in our analysis of music, mapping musical events to events in the world and asking whether certain key relations among musical events are preserved by the map. The analogy is not coincidental, since we take music semantics to have the same general structure as other inferential systems in perception.

6.3 Comparisons

In this section, we briefly compare our music semantics to more standard varieties of semantics: standard logical semantics; and the iconic semantics that were developed for certain aspects of sign language, and for pictures. We argue that music semantics is very different from logical semantics, but more comparable to iconic semantics.

6.3.1 Differences between music semantics and logical semantics

In order to compare music semantics to logical semantics, we define a non-standard but particularly simple logic that shares some properties with our music semantics; the main differences will become easier to grasp within this shared background. In a nutshell, this logical semantics is defined for a language made solely of atomic propositions and conjunctions. Since the only way to combine propositions is by way of conjunction, we don't need an explicit conjunction sign and thus we will solely obtain sentences of the form: p_i, $p_i p_k$, $p_i p_k p_r$, etc.

(38) A toy logical semantics
a. Syntax
(i) Atomic propositions: for every $i \geq 0$, p_i is an atomic proposition
(ii) If F and G are propositions, $F\ G$ is proposition

b. Semantics

Let I be a function such that for every $i \geq 0$, $I(p_i)$ is a set of events.
(i) For any propositional letter p_i, p_i is true of event e just in case e is in $I(p_i)$
(ii) If p_i is an atomic proposition and F is a proposition (whether atomid or not), $p_i\ F$ is true of event e just in case p_i is true of e and F is true of e.

(39) Examples
$I(p_1) = \{e, e'\}$
$I(p_2) = \{e, e', e''\}$
$I(p_3) = \{e', e''\}$

a. For any event f, $p_2\ p_3$ is true of f iff p_2 is true of f and p_3 is true of f, iff f is in {e, e', e"} and {e', e"}, iff f is in {e', e"}, iff f = e' or f = e"

b. For any event f, $p_3\ p_2$ is true of f iff p_3 is true of f and p_2 is true of f, iff f = e' or f = e" (by a.)

c. For any event f, $p_1\ p_2\ p_3$ is true of f iff p_1 is true of f and $p_2\ p_3$ is true of f, iff f is in {e, e'} and f is in {e', e"} (by a.), iff f = e'

Our rules are designed in such a way that a string is always semantically analyzed from beginning to end. Nothing deep hinges on this: given our conjunctive semantics, whether a string $p_i\ p_k\ p_m$ is analyzed as as $[p_i\ [p_k\ p_m]]$ (as we do) or as $[[p_i\ p_k]\ p_m]$ won't affect the truth conditions, since the end result will just be the conjunction p_i *and* p_k *and* p_m.

One could think of $p_i p_k\ p_m$ as a series of musical events, which may be true some events such as e, e', e", etc. But the similarities with our music semantics end here. First, there are no counterparts of our preservation principles (Loudness, Harmonic stability) in this case; rather, we stipulate that a proposition is true of certain events, without trying to derive from the shape of the propositional letter what events it is true of. Second, when we combine two propositions of our toy logical semantics, the order in which they are combined is irrelevant to the meaning of the result, as seen in (39)a-b; this is very different from the case of music semantics, in which we took the sequence of musical events to be dynamic representations of real world events, with the result that the order in which the musical events appear crucially affects the resulting meaning.

6.3.2 Connection with iconic semantics and picture semantics

If one wishes to find a better point of comparison for music semantics, one should consider dynamic visual representations such as films or iconic gestures and signs. We discussed at the outset the relevance for music semantics of Heider and Simmel's abstract animations, in which animated shapes took the character of agentive entities depending on their movements. But simpler cases of dynamic pictorial representations – even without a notion of agency – can be profitably compared to music semantics.

We start from a simple iconic example from American Sign Language. Sign languages notoriously have the same grammatical and logical structure as spoken languages, but *in addition* they can make use of rich iconic resources, illustrated here with the verb *GROW*. It can be realized in a variety of ways, six of which were tested in (41).

(40) POSS-1 GROUP GROW.
'My group has been growing.' (8, 263; 264) (Schlenker et al. 2013)

(41) Representation of *GROW*

	Narrow endpoints	**Medium endpoints**	**Broad endpoints**
Slow movement	small amount, slowly 	medium amount, slowly 	large amount, slowly 
Fast movement	small amount, quickly	medium amount, quickly	large amount, quickly

The sign for *GROW* in (40) starts out with the two hands forming a sphere, with the closed fist of the right hand inside the hemisphere formed by the left hand; the two hemispheres then move away from each other on a horizontal plane (simultaneously, the configuration of the right hand changes from closed to open position). The signer varied two main parameters in (41): the distance between the

endpoints; and the speed with which they were reached.[19] All variants are acceptable, but yield different meanings, as shown in (41). Intuitively, there is a mapping between the physical properties of the sign and the event denoted: the broader the endpoints, the larger the final size of the group; the more rapid the movement, the quicker the growth process.

Formally, two properties of the sign are preserved by semantic interpretation, as stated in (42).

(42) Preservation requirements on the interpretation of *GROW*
Let $GROW_i$ and $GROW_k$ be two realizations of the sign *GROW*, and let e_i and e_k be two events of growth that are in the extension of $GROW_i$ and $GROW_k$ respectively. Then:

a. Breadth condition
If the end points of $GROW_i$ are less distant than those of $GROW_k$, then the endpoint of the growth in e_i should be smaller than that of the growth in e_k.

b. Speed condition
If sign $GROW_i$ is realized less fast than $GROW_k$, the growth in e_i should be slower than the growth in e_k.

As can be seen, the preservation conditions are formally relatively similar to those we posited in the Loudness and the Harmonic stability condition in our toy model. Still, there is one important difference. In our sign language example, the iconic conditions *enrich* a verbal meaning. *GROW* is a verb, and thus like the English verb *grow* it should be a predicate – within a ('Davidsonian') semantics, it is true of events of growth. Because this verb *also* has an iconic life, its meaning is enriched with by the preservation requirements in (42). By contrast, we took musical events to be auditory traces or pictures of some events, and thus we defined a semantics in which a musical event denotes a real world event rather than a set of them, as would befit a predicate.

Greenberg 2013 defines a formal semantics for pictures, which unlike the case of ASL *GROW* is *purely* iconic. To obtain a visual analogue of music semantics, one should investigate the semantics of (possibly abstract) animations, which unlike pictures have a dynamic component.[20]

7 The Syntax/Semantics Interface

Following Lerdahl's and Jackendoff's ground-breaking work (1983), much attention has been devoted in the last thirty years to the *structure* of music. While one could take a music semantics to exist alongside a music syntax without directly interacting with it, we will ask in this section *how much of music structure can be derived from principles of music semantics*. We will argue that Lerdahl and Jackendoff's 'grouping structures' are semantic from the get-go (because they are explicitly 'lifted' from Gestalt principles of perception), and that the time-span reduction structures that are built out of them can naturally be re-interpreted in semantic terms as well.

7.1 Lerdahl and Jackendoff's hierarchical structures

Lerdahl and Jackendoff posit four levels of structure, summarized as follows in Lerdahl 2001:

GTTM proposes four types of hierarchical structure simultaneously associated with a musical surface. Grouping structure describes the listener's segmentation of the music into units such as motives, phrases, and sections. Metrical structure assigns a hierarchy of strong and weak beats. Time-span reduction, the primary link between rhythm and pitch, establishes the relative structural importance of events within the rhythmic units of a piece. Prolongational reduction develops a second hierarchy of events in terms of perceived patterns of tension and relaxation.

The four hierarchies integrate (leaving aside feedback effects) as in Figure 1.2. From the grouping and metrical structures the listener forms the time-span segmentation over which the dominating—subordinating relationships of time—span reduction take place; and from the time-span reduction the listener projects the tensing—relaxing hierarchy of prolongational reduction. (...)

[19] The paradigm was not fully minimal, in the sense that further aspects of the sign tended to be modified as well.

[20] Abstract animations that were designed to complement musical pieces would be particularly interesting to investigate in this connection. A nice example is offered by Mary Ellen Bute's *Tarantella* [http://www.dailymotion.com/video/xgd0wn_tarantella-1940_shortfilms], an abstract animation that was conceived in conjunction with piano music by Edwin Gerschefski. One could explore in future work the ways in which the music and the visual animation converge on a single semantic effect or not.

(43) Figure 1.2. from Lerdahl 2001

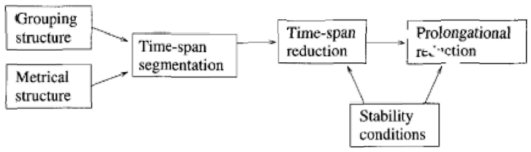


FIGURE 1.2 A flow chart of GTTM's components.

Some of Lerdahl and Jackendoff's structures have been analyzed in terms of a generative syntax, as in Pesetsky and Katz 2009 for prolongational reductions. Lerdahl and Jackendoff's own discussion departs in two respects from a 'generative syntax' analysis.
(i) First, they take their structures to be based on parsing rather than generation, and to be rely heavily on preference principles rather than categorical principles of well-formedness.
(ii) Second, Lerdahl and Jackendoff take some of their own structures to be based in perception and to follow from very general Gestalt principles.
(i) may or may not be essential, for one might present the same system in terms of parsing or generation, as Pesetsky and Katz argue. But (ii) is very essential for present purposes, as it suggests that *the rules that provide structure to musical form are rules of perception designed to capture the structure of the represented events*.

7.2 Grouping structure and event mereology

We will now suggest that Grouping structures are best seen as originating in the mereological structure of events, i.e. the part-of structure (sometimes called 'partology') of events. More specifically, we take Grouping structure to derive from the fact that the auditory traces of (real word) events are organized in a way that reflects the structure of these events. In some cases this gives rise to a tree-like structure, but for reasons that are very different from what we find in human language.

We will proceed in three steps. First, we will note that it uncontroversial that events come with a part-of structure (large events are made of smaller events), and that with additional assumptions a tree-like structure is obtained. Second, we will argue that the result is a more flexible theory of music structure than a syntactic tree structure would yield, because in some cases it allows for overlap among groups, or discontinuous groups. Third, we will refer to literature on event perception to suggests that these are indeed perceived as structured.

7.2.1 Event mereology

It is uncontroversial that events have a part-of structure, with large events being made of smaller events. Still, the part-of structure is very weak, and thus further assumptions are needed to obtain tree-like structures.

We will start from the simple part-of structure given in (44); it has in particular the consequence that if an event e has parts, then *their* parts are also parts of e (Transitivity).

(44) Part-of structure in mereology (e.g. Varzi 2015)
The part-of relation P is defined by the following requirements (for all x and y):
a. Reflexivity
For all x, Pxx
b. Transitivity
For all x, y, if Pxy and Pyz, then Pxz
c. Antisymmetry
If Pxy and Pyx, x = y

The notion of 'proper part' follows from that of a part: x is a proper part of y iff x is a part of y and x and y are not identical. For simplicity, we will further assume that every event is made of atomic events, i.e. events that do not themselves have proper parts, as defined in (45).

(45) Atoms (e.g. Varzi 2015)
a. Definition: x is an atom iff x has no proper part.
b. Atomicity:
For all x, x has a part which is an atom

(46) Assumption: every event is made of atomic events.

Assuming that this structure applies to events, we can define a partially ordered structure in which an element immediately dominates its immediate proper parts, and restrict attention to graphs that lead to atoms. Among all the structures of this sort, we will obtain tree structures as special cases – but further assumptions are needed to get there.

First, it makes sense to assume that atomic events are ordered in time, as stated in (47).

(47) If x and y are atomic events, either x < y or y < x, where < is a temporal ordering.

We henceforth use the list of its atoms to name an event. Let us now look at all the possible decompositions of an event abc, with a < b < c.

(48) Possible decompositions of abc
a. abc –> a, b, c

b. abc –> ab, c
ab –> a, b

c. abc –> a, bc
bc –> b, c

d. abc –> **ac**, b
ac –> a, c

e. abc –> **ab, bc**
ab –> a, b
bc –> b, c

Things will be a bit more legible if we omit 'trivial' decompositions, namely ones that involve events with just two atomic parts (since these can be decomposed in just one way). This leads to the representations in (49).

(49) Possible decompositions of abc - simplified notation
a. abc –> a, b, c

b. abc –> ab, c

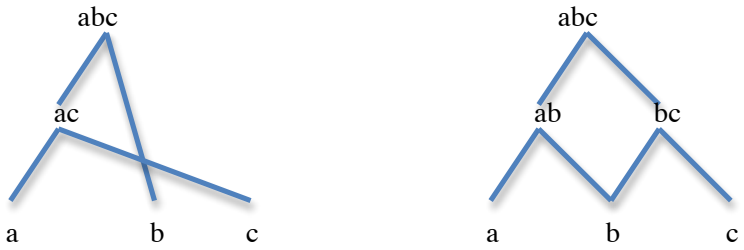
c. abc –> a, bc

d. abc –> **ac**, b

e. abc –> **ab, bc**

Now it can immediately be seen that (49)a, b, c correspond to 'standard' derivation trees that could be obtained from a context-free grammar. But (49)d, e have an unusual shape:

(50) a. b.



• The situation in (50)a violates the assumption that 'constituents are not discontinuous' (a standard but not universal assumption in linguistics, see e.g. McCawley 1982 for exceptions). In standard syntax, it is normally prohibited by the assumption that in a context-free rule of the form M –> $D_1 \ldots D_n$, the output elements $D_1 \ldots D_n$ are temporally ordered with $D_1 < \ldots < D_n$, with a requirement that if $D_i < D_k$, then all the terminal nodes dominated by D_i precede all the terminal nodes dominated by D_k (see Kracht 2003 p. 46); precisely this condition fails in (50)a, as we can neither have ac < b nor b < ac.

• The situation in (50)b violates the assumption that a terminal node is the output of a single context-free rule, so that 'multi-dominance' is prohibited (this prohibition was reconsidered in syntax in theories of 'multidominance' (e.g. de Vries 2013)).

Can these structures be blocked in a natural way if we take them to reflect event structure? We believe that they can be.

• Consider first (50)b. It is an uneconomical event decomposition, because we could remove a branch above b (thus attributing b exclusive to the left-hand or to the right-hand node that dominates it)

without affecting the set of atomic elements that constitute the whole. This condition of economy can be enforced by (51), which prohibits overlap among events unless one is contained within the other.

(51) Minimal part-of structures
A part-of structure is minimal if whenever x is part of y and x is part of z, y is part of z or z is part of y.

This condition is of course violated by (50)b: b is part of ab and of bc, but neither is part of the other.

We take this minimality condition to be a principle of optimal event perception, but one that should have exceptions. These could be of two sorts:
(i) overlap: cases in which there is a reason to think that the represented (real world) events are best decomposed in a non-economical fashion, with a part which is common to both (for instance because there is a smooth transition between two events[21]);
(ii) occlusion: cases in which there is a reason to think that two distinct events share the same auditory trace.
We come back in Section 7.2.2 to exceptions of both sorts.

• Consider now (50)a. It leads one to posit that an event has a discontinuous auditory trace. Two assumptions are needed to prohibit this case.
(i) The first assumption, which makes much intuitive sense, is that real-world events are normally connected. But this measure is not enough. Consider an analogous case in the visual domain. It makes sense to posit that both objects and events satisfy a condition of spatial or temporal connectedness. Still, due to occlusion, there are numerous objects and events that we *see* as disconnected, even when our cognitive system is able to take occlusion into account and to posit single a single underlying object or event despite the apparently disconnected nature of the percept.
(ii) In order to prohibit structures such as (50)a, then, we must also posit that cases of auditory occlusion don't occur. This makes much sense in some standard situations: if you are in the middle of a conversation while a car passes by, it will rarely happen than the background noise is so strong as to occlude the conversation, or conversely.

In this case as well, we predict that there should be exceptions, of two types.
(i') There could be cases in which it makes sense to assume that the connectedness condition fails to apply to real-world events.
(ii') There could also be cases in which the connectedness condition does apply to real world events, but not to their auditory traces, in particular due to cases of occlusion.

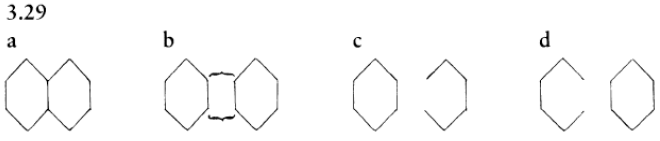
7.2.2 Exceptions

Lerdahl and Jackendoff 1983 emphasize that cases such as (50)b arise in music. Since they take grouping structure to result from principles of perception rather than from syntactic rules, they do not take these 'exceptions' to refute their account; on the contrary, they explain these exceptions appealing to analogous cases in visual perception. Furthermore, the exceptions they list are of the two types we announced above: in case so overlap, the denoted events are construed as sharing a part; in cases of occlusion, the auditory trace of an event occludes that of another event.

- *Overlap*

Lerdahl and Jackendoff 1983 illustrate visual overlap by the case in which a single line serves as the boundary between two objects, and is thus best seen as belonging to both, as in (52)a, which is preferably analyzed as (52)b rather than as (52)c-d. In our terms, this is a case in which the optimal mereological decomposition of the underlying object should not be minimal – although an alternative possibility is that we are dealing with two different lines that have a unique visual trace.

(52) Lerdahl and Jackendoff's visual analogue of overlap (Lerdahl and Jackendoff 1983 p. 59)

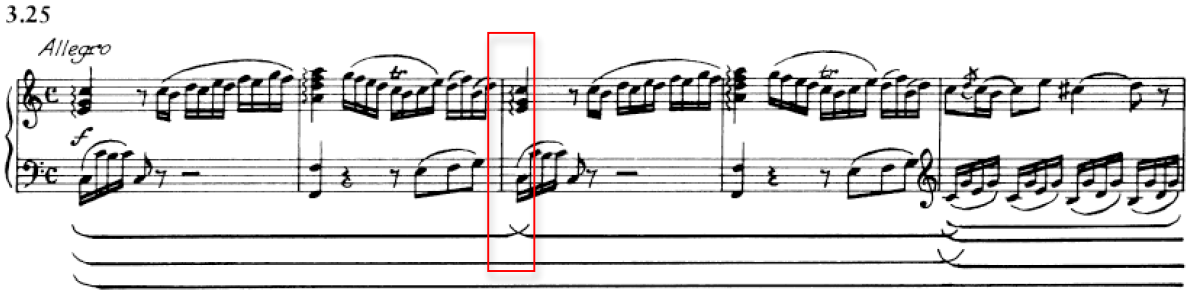


Lerdahl and Jackendoff 1983 cite the very beginning of Mozart's K. 279 sonata as an example of auditory overlap, as seen in (53). The I chord at the beginning of bar 3 seems to both conclude the first group and initiate the second, hence it can be taken as the trace of an event that plays a dual role as the end of one event and at the beginning of another. Alternatively, and less plausibly perhaps, this

[21] Cases of modulation might be of this type.

could be a case in which two distinct events have the same auditory trace (this is precisely the uncertainty we had in our discussion of the visual example in (52)).[22]

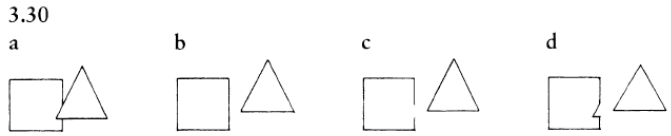
(53) An example of overlap: the beginning of Mozart's K. 279 sonata (Lerdahl and Jackendoff 1983 p. 56)
https://www.youtube.com/watch?v=d26zRUWKc08



❑ *Occlusion*

The second case involves part of an object occluding another object, as in (54). Here the most natural interpretation of (54)a is as (54)b, which involves occlusion, rather than as (54)c-d, which don't.

(54) Lerdahl and Jackendoff's visual analogue of elision (Lerdahl and Jackendoff p. 59)



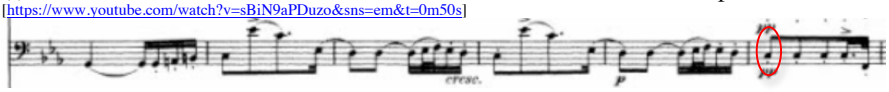
In music, this case is illustrated by what Lerdahl and Jackendoff call 'elision'. Their description (as well as the visual they draw) makes clear that these are really cases of auditory occlusion, as in their discussion of the beginning of the allegro of Haydn's Symphony 104. As they write:

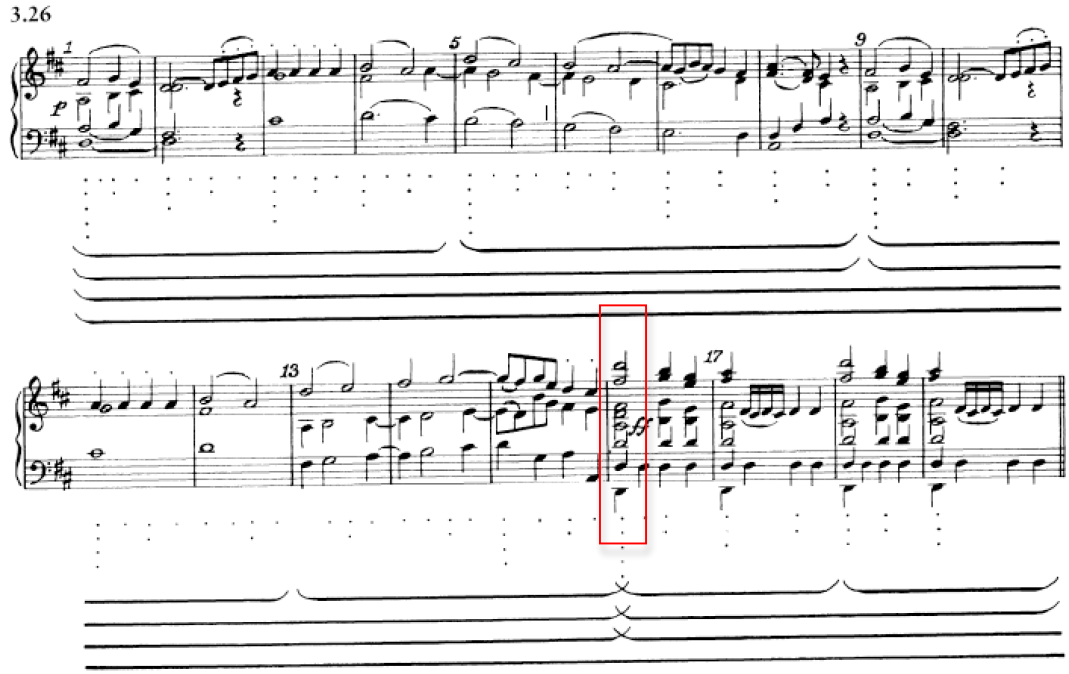
One's sense is not that the downbeat of measure 16 is shared (...) ; a more accurate description of the intuition is that the last event [of the first group] is elided by the fortissimo.

(55) An example of elision: the beginning of the allegro of Haydn's Symphony 104 (Lerdahl and Jackendoff 1983 p. 57)
https://www.youtube.com/watch?v=OitPLIowJ70&t=2m14s

[22] A radical version of the same effect can be found in (i), where the cello initially carries the melody, and becomes an accompaniment in the last bar. The circled note is both the conclusion of the melody and the beginning of the accompaniment.

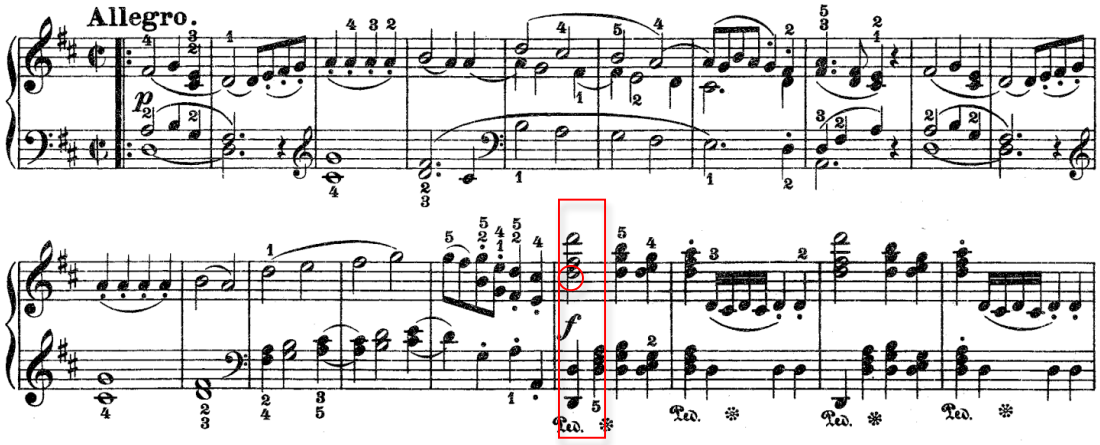
(i) Measures 16-21 of Schubert's Piano Trio in E-flat D929 - Cello part
[https://www.youtube.com/watch?v=sBiN9aPDuzo&sns=em&t=0m50s]





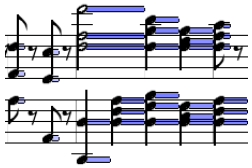
To see Lerdahl and Jackendoff's observation in action, we consider August Horn's piano transcription of the relevant bars of Haydn's Symphony 104. It differs in particular from Lerdahl and Jackendoff's transcription in that the expected D (shown circled in (56)a) has been re-introduced. One can then compare this version with one from which this D is removed, and with a third one from which all D's in that chord except the highest one have been removed, as in (57)c. In the latter two cases, it's probably easy to hear the 'missing' D as being present but 'occluded' by the rest of the chord.

(56) Horn's piano transcription of the beginning of the allegro of Haydn's Symphony 104

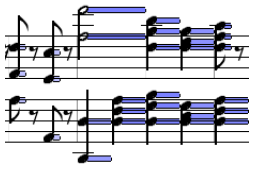


(57) Three versions of the forte chord in Horn's transcription (credit: A. Bonetto)

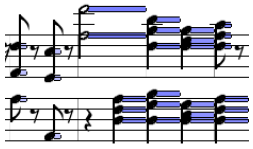
a. Horn's version (= (56))
https://soundcloud.com/philippeschlenker/haydn104v1-all-ds



b. Removing the circled D in (56)
https://soundcloud.com/philippeschlenker/haydn104v2-no-penultimate-d

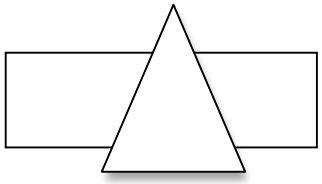


c. Removing all D's except the highest one in (56)
https://soundcloud.com/philippeschlenker/haydn104v3-only-highest-d



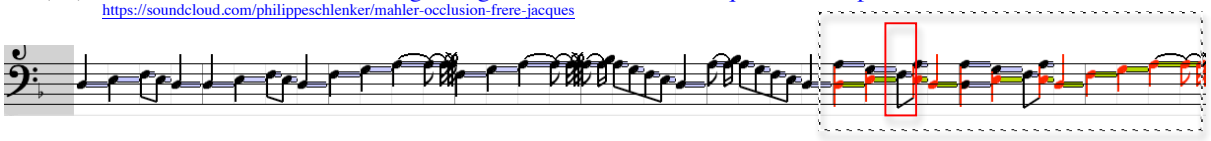
This case of occlusion is rather special in that it happens at the end of a group. In the visual domain, an object can appear as discontinuous because it is occluded by another, as shown in (58), which is but a modification of Lerdahl and Jackendoff's own examples in (54).

(58)



Can we find such cases of occlusion in music? They could be expected to involve two voices, since one of the events must be auditorily occluded by another event that co-occurs with it. This frequently happens on the piano when a key is 'shared' between the two hands. A simple modification of Mahler's Frère Jacques can make the point, as in (59). Since the boxed F at the end is shared between two voices, it is clear that if this note is attributed to one voice the other will be discontinuous, and thus we must have a case of partial occlusion of one voice by the other.

(59) A modified version of the beginning of Mahler's Frère Jacques, on the piano, with occlusion
https://soundcloud.com/philippeschlenker/mahler-occlusion-frere-jacques



The end of the theme (dotted boxed part) has been modified, and as a result the boxed F is shared between the end of the theme and the second occurrence of the beginning of the same theme.

7.2.3 Sequencing events

Jackendoff 2009 notes that there are tree-like structures outside of language, and he gives the example of actions, which may be structured ways without thereby having a linguistic representation; we come back to his examples below. For the moment, let us note that Zacks et al. 2001 provide experimental evidence that subjects sequence events (by way of videos) in a hierarchical fashion. Specifically,

> participants were asked to segment everyday activities while watching them. Each participant segmented each activity twice, once under fine-unit and once under coarse-unit coding instructions. Within individuals, the boundaries of the coarse units tended to be closer to the boundaries of fine units than predicted by chance. This alignment effect was mediated by the familiarity of the activity and was more pronounced when participants described the activity while segmenting it than when they only performed the hierarchically organized schemata for recurring events.

Finding that descriptions of the events access the same hierarchies, they conclude that "the data argue strongly for the hierarchical bias hypothesis: Observers' perception of event structure is biased by the influence of hierarchically organized schemata for recurring events."

Of course the fact that the boundaries of coarsely-individuated groups ('large groups') should correspond to the boundaries of some finely-individuated groups ('small groups') is a pre-requisite for a hierarchical event structure, but more specific results would be needed to know what constraints on events we should initially posit. Two questions would be particularly interesting for music semantics:

• Are there systematic cases in which two events may overlap, something which is crucial to our understanding of Lerdahl and Jackendoff's cases of overlap?

• Are there natural cases of discontinuous events? If the answer is positive, one might add a new category of discontinuous groups in music: besides the cases we discussed on the basis of auditory occlusion, there could be cases in which a musical group is analyzed as being discontinuous because the event it is the auditory trace of is itself discontinuous.

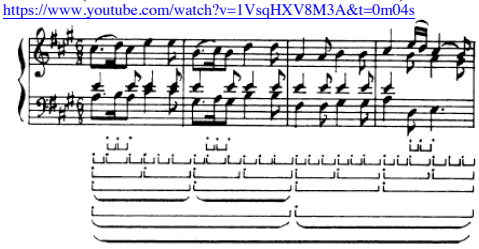
In sum, in several cases grouping structure departs from a simple tree structure, in ways that can be explained if musical groups are perceived as the auditory traces of events, whose mereological structure is reflected on the musical surface. In particular, there are cases of overlap in which a part is best seen as belonging to two events, cases of occlusion in which the auditory trace of one event occludes that of another event; the latter phenomenon may occur at the end of events, but also within events, in which case groups may be discontinuous.

7.3 Time-span reductions and headed events

It is uncontroversial that Western classical music has a metrical structure that yields an alternation of strong and weak beats. Lerdahl and Jackendoff 1983 analyze it with rules that are very similar to those used in metrical phonology. Thus at small levels musical events are structurally organized by metrical structure, while at larger levels they are organized by grouping structure. But Lerdahl and Jackendoff argue that grouping structure is not enough; rather, some subevents are conceived as more important than others. Formally, they propose that their tree structures should be seen as being *headed*: in each natural unit, one musical event is more important than the others and is its 'head'. In a nutshell, heads are events that are rhythmically more prominent and/or harmonically more stable. In their words,

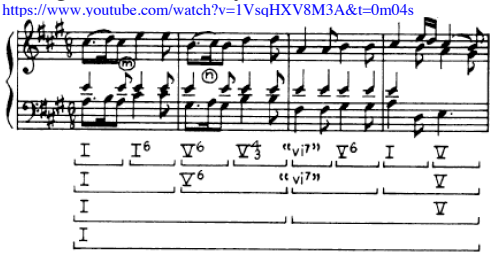
at the most local levels, the metrical component marks off the music into beats of equal time-spans; at larger levels, the grouping component divides the piece into motives, subphrase groups, phrases, periods, theme groups, and sections. Thus it becomes possible to convert a combined metrical and grouping analysis into a time-span segmentation, as diagrammed for the beginning of [Mozart's] K. 331 in 5.11.

(60) Metrical structure [segments] and grouping structure [brackets] for the beginning of Mozart's K. 331 piano sonata (Lerdahl and Jackendoff 1983)
https://www.youtube.com/watch?v=1VsqHXV8M3A&t=0m04s



The next step in the construction of time-span reductions is the selection of a head in each group, as is illustrated in (61).

(61) Time-span reduction obtained from (61) by selecting in each the musical event which is metrically strongest/harmonically most stable (Lerdahl and Jackendoff 1983)
https://www.youtube.com/watch?v=1VsqHXV8M3A&t=0m04s



As Lerdahl and Jackendoff write (p. 120), "in the span covering measure 2, the V^6 is chosen over the V^4_3, and proceeds for consideration in the span covering measures 1-2.; here it is less stable than the opening I, so it does not proceed to the next larger span; and so forth. As a result of this procedure, a particular time-span level produces a particular reductional level [the sequence of heads of the tirne-spans at that level)."

It remains to ask whether the headed nature of time-spans should be taken as primitive, or follows from a more general strategy of event perception. Jackendoff 2009 argues that there are

headed structures outside of music and language[23], in particular in the domain of complex action. As a examples, he discusses making coffee in an electric coffee maker:

this is a recursive headed hierarchy. The basic element in the tree is an event consisting of a Head (the main action), with an optional Preparation (things that have to be done before the Head can be begun) and an optional Coda (things that are done to restore the status quo ante). For instance, consider the constituent "put water in machine" in Figure 2. The Head consists of actually pouring water into the machine from the pot. But in order to do this, one must first measure water into the pot—the Preparation—which in turn is organized into Preparation plus Head, and each of these has further organization. And once one has poured the water into the machine, one must replace the empty pot in the machine—the Coda.

(62) Making coffee in an electric coffeemaker according to Jackendoff 2009

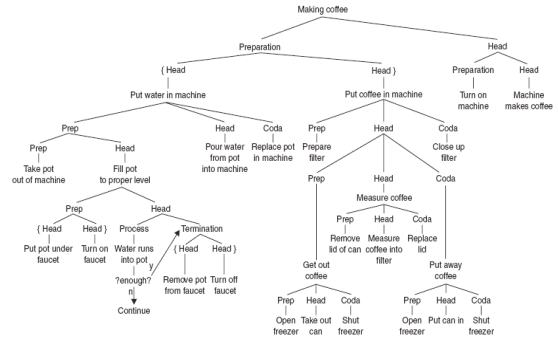


FIGURE 2. Structure of the complex action of making coffee.

Now Jackendoff only mentions the example of complex action to suggest that headed structures exist outside of language, and thus that their existence in music does not argue for a 'linguistic' approach to musical structure. He does speculate that the "integration and execution of complex action might be a strong candidate for a more general, evolutionarily older function that could be appropriated by both language and music", but he does not otherwise use the structure of action as a way to explain musical structure.

From the present perspective, however, a natural question is whether we could explain the headed nature of time spans as reflecting the headed nature of the denoted events. We conjecture that this is indeed the case, and specifically:
(i) that real world events are perceived not just as structured but also as headed (at least in some cases), and that
(ii) that considerations of energy (comparable to rhythmic strength) and of stability (comparable to harmonic stability) both play a role in selecting the head of an event.
While this is pure speculation at this point, we would like to discuss two suggestive examples.

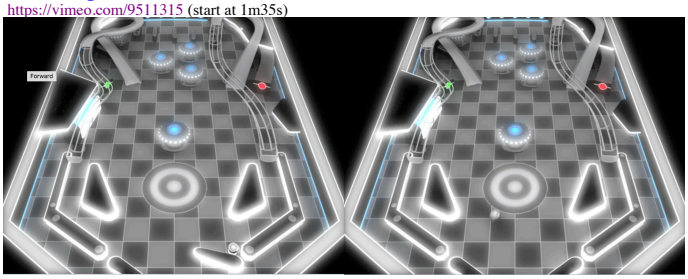
• First, consider a simplified dynamic representation of a person walking, as in (63). We submit that if one were to sequence the walk into events and subevents, one would find that moments at which the foot touches the ground delimit events, but in addition that these are the most important events in each cycle – the 'head' of the relevant event, in terms of the present discussion. These are clearly points at which impulses of energy are given, somehow like points of metrical strength in music. (Anecdotally but probably not coincidentally, if one walks or performs some physical task while listening to music, one tends to synchronize to the strong beats.)

(63) Person walking
https://www.youtube.com/watch?v=ZPI7_oVNB24

[23] In language, it is uncontroversial that groups are headed: Verb Phrases of different natures (intransitive construction, transitive construction, ditransitive constructions) share important properties by virtue of being headed by a verb. In music, Jackendoff 2009 discusses prolongational structures rather than time-span structures, but both have a headed structure in the analysis of Lerdahl and Jackendoff 2083.

• Second, consider a simplified pinball game. One could ask subjects to sequence the movement of the ball. Our guess is that grouping would follow standard Gestalt principles of 'good form', but that in addition subjects might be able to select subevents as being more important than others, with impulses of energy and possibly points of stability playing a distinguished role. This remains to be investigated. (The images below are from pinball animations created for a musical – with the effect that one can in principle investigate the coordination between time-span heads and 'strong' events.)

(64) Pinball game
https://vimeo.com/9511315 (start at 1m35s)



7.4 Structural interpretive rules?

At this point, we have argued that time-span structures should be taken to derive from principles of event perception. Still, one could also start from musical structure and ask how headed time-span groups should be semantically interpreted. If we had a semantics for elementary musical events (something we have not fully developed in this piece), we could attempt to extend it to larger structures by way of the rule in (65), where [[•]] is the interpretation function, which assigns to a musical event • the set of its possible denotations, and where + is used to represent event summation.

(65) Let H and N be two musical constituents, with H a head and N a non-head (in the time-span tree representation of Lerdahl of Jackendoff 1983).
a. [[H N]] = {s+s': s is an event in [[H]] and s' is an event in [[N]] and s immediately precedes s' s is more important than s'}
b. [[N H]] = {s+s': s is an event in [[N]] and s' is an event in [[H]] and s simmeidately precedes s' and s' is more important than s}

In a nutshell, this rule interprets subtrees of the form H N, where H is the head of the larger constituent, and takes it to denote the set of sequences of events s+s', where s is a possible denotation of H, s' is a possible denotation of N, the temporal ordering of s and s' corresponds to that of H and N, and crucially s is more 'important' than s'. The notion of importance would of course need to be clarified, and we conjecture that notions of energy and stability would play a role in it.

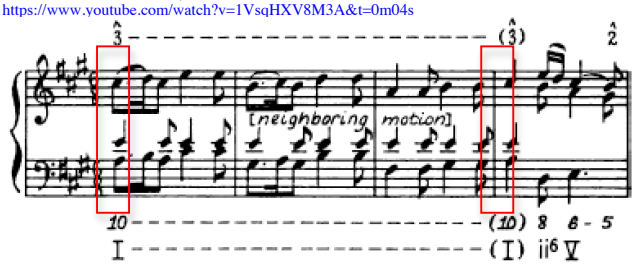
7.5 A note on prolongational reductions

Lerdahl and Jackendoff 1983 and Lerdahl 2001 take another notion of structure, prolongational reductions, to play a central role in music perception.[24] Specifically, prolongational reductions provide a hierarchy of events "in terms of perceived patterns of tension and relaxation" (Lerdahl 2001, cited above). To make things concrete, consider again the time-span structure in (61). Lerdahl and Jackendoff argue that it is incapable of representing the intuitive patterns of tension and relaxation represented in (66):

One might say that the phrase begins in relative repose, increases in tension (second half of measure 1 to the downbeat of measure 3), stretches the tension in a kind of dynamic reversal to the opening (downbeat of measure 3 to downbeat of measure 4), and then relaxes the tension (the rest of measure 4). It would be highly desirable for a reduction to express this kind of musical ebb and flow. Time-span reduction cannot do this, not only because in such cases as this it derives a sequence of events incompatible with such an interpretation ([(61)] as opposed to [(66)]). but because the kind of information it conveys, while essential, is couched in completely different terms. It says that particular pitch—events are heard in relation to a particular beat, within a particular group, but it says nothing about how music flows across these segments. (Lerdahl and Jackendoff 1983 p. 122)

[24]Pesetsky and Katz 2009 take prolongational reductions to be central to their 'identity thesis' for music and language. For them, time span reductions share properties with prosodic structure in phonology, whereas prolongational reductions play the role of (and shares properties with) syntactic structure. They further suggest that prolongational reduction need not be taken to be derivative from time span reductions, as argued by Lerdahl and Jackendoff 1983 and Lerdahl 2001.

(66) Prolongation of the initial I chord at the beginning of of Mozart's K. 331 piano sonata (Lerdahl and Jackendoff 1983)
https://www.youtube.com/watch?v=1VsqHXV8M3A&t=0m04s



In the time-span structure in (61), the last bar forms a group, but it is headed by the V chord, which is harmonically essential (as it marks a half-cadence). As a result, the I chord at the beginning of the last bar plays a subordinate role. But intuitively corresponds to the end of a tensing and relaxing motion that started at on the same I chord, but at the beginning of bar 1. In Lerdahl and Jackendoff's analysis, prolongational structures are derived top-down from time-span structures in such a way that subordinate time-span events can be 'promoted' to a higher hierarchical level if they play a key role in patterns of tension and relaxation.[25]

From the present perspective, two main questions arise about prolongational reductions.

• First, could they have a counterpart in other areas of perception? In particular, if we could find visual scenes with (i) 'headed' events' (in order to have a counterpart of time-span structures), and (ii) a natural notion of tension (e.g. in terms of more or less stable physical situation), could we also elicit intuitions about an equivalent of prolongational structures? This is what one would expect if the difference these two kinds of structures derives from the kind of semantic information they seek to capture (event mereology for time-span structures, properties of the path of events in a certain space for prologational structures).

• Second, could one investigate structures that capture other salient properties of musical events, such as melodic line or loudness? A key insight of Lerdahl and Jackendoff's analysis of prolongational structure is that two musical events that are not linearly contiguous may have a direct structural connection, as is the case of the two highlighted I chords in (66). It would be interesting to determine whether other properties besides tension/relaxation give rise to structural representations of this sort which depart from time-span reductions (and if so, how these structural representations can be derived).

8 Pragmatics

At this point we have been solely concerned with music syntax and semantics. Let us say a few words about what a music pragmatics could look like.

In linguistics, 'pragmatics' usually makes reference to aspects of language use that do not just derive from its intrinsic structure, but also from properties of communicative rationality. Here we will focus on three issues: How is information structured by the musical narrator? What are the various levels at which intentional effects can be found in music? Are there musical equivalents of dialogues? In each case, we only aim to formulate the main questions, leaving it to future research to seriously address them.

8.1 Information structure

As we mentioned at the outset, information may be structured even in a system which lacks as semantics, such as the syllable sequences we discussed at the outset (as in (11): [la lu] [la lu] [la LI] [la lu]). One would expect such effects to hold in music as well, but there are now two reasons for which this may be the case:

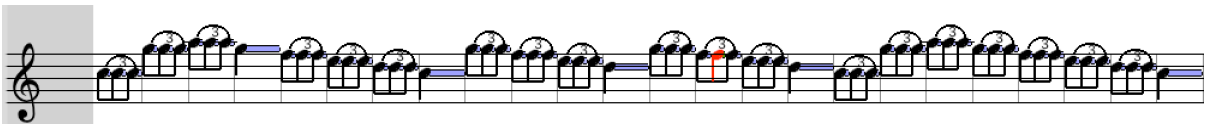
(i) it could be that just as in syllable sequences the mere form of music conveys information, and is structured for this reason;

(ii) but in addition, there might be cases in which musical information is structured due to its semantic content.

[25] Pesetsky and Katz 2009 argue instead for mapping rules which do not give precedence to time-span reductions.

Case (i) might be exemplified in the following modification of Mozart's *Ah vous dirai-je maman*: triplets have been introduced to ensure that notes are repeated on weak beats, and of course the theme involves repetitions as well.[26] Now the highlighted F in (67)a conveys doubly old information: first, because it appears in the second position of a series of notes that are predictably repeated; second, because the three-bar phrase it belongs to it itself the repetition of the preceding phrase. As a result, playing this note louder than the immediately preceding F is odd, as the highlighted note is in a weak beat and conveys old information. By contrast, if this F is replaced with an A or a D, as in (67)b, the result is more natural, presumably because the note is now unexpected and provides new information.

(67) Modification of Ah vous dirai-je maman, with triplets[27]
a. Simple version with triplets
https://soundcloud.com/philippeschlenker/mozart-ah_vous-focus-triplets

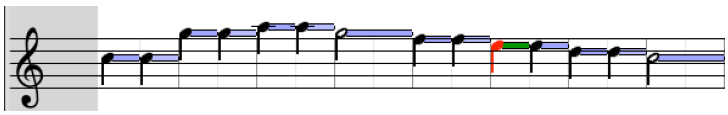


b. Modified version with with an A replacing the highlighted F in a.
https://soundcloud.com/philippeschlenker/mozart-ah_vous-focus

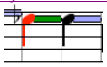
c. Modified version with with an A replacing the highlighted F in a.
https://soundcloud.com/philippeschlenker/mozart-ah_vous-focus-1

A schematic attempt to illustrate a possible instance of Case (ii) is given in (68). Here we contrast a normal, major version of Ah vous dirai-je maman with a version in which the second phrase is made minor by turning an E into an Eb. As a result, this Eb conveys important harmonic information. If the first Eb in (68)b is accented, the result sounds rather normal, presumably because of the importance of its informational content. But if the homologue E is similarly accented in (68)a, the result is a bit odd, because nothing justifies highlighting this note.

(68) Modification of Ah vous dirai-je maman, adding an accent on the highlighted note
a. Simple version, major: an accent on the highlighted E is a bit odd
https://soundcloud.com/philippeschlenker/mozart-ah-vous-dirai-je-maman-melodic-base-minor-loud-control



b. Modified version, with an Eb replacing the highilighted E, thus making the second phrase minor: an accent on the highlighted Eb is more natural tha one on the highlighted E in a.
https://soundcloud.com/philippeschlenker/mozart-ah-vous-dirai-je-maman-melodic-base-minor-loud



Needless to say, these examples would need to be studied much more systematically before it can be asserted that accent has the informational function we claim.[28] We mention this possibility because it highlights one role of pragmatics in music, involving information structure.

8.2 Levels of intentionality

More generally, pragmatics is based on the premise that the speaker is an intentional agent and obeys some principles of rationality and specifically of cooperative information exchange. However, there are further intentional entities that may play a role in music semantics, and it thus worth distinguishing the various levels at which intentional effects could arise. These distinctions could matter in the analysis of musical pieces.

---

[26] Thanks to A. Bonetto for suggesting that we consider a version with triplets.
[27] The sound examples were produced as followed: Bonetto produced (67)c on an electronic piano. (67)a-b were produced from the recorded version of (67)c via manipulations (with the software GarageBand).
[28] Jonah Katz and Emmanuel Chemla are doing experimental work on this topic.

• First, we took musical voices to be associated with objects, which may be intentional or not. In opera, they are typically associated with individuals – and thus the re-assertion we discussed in connection with Mozart's 'Rispondimi!' in Don Giovanni (in (25) above) is interpreted as a re-assertion on the Commendatore's part. Intentional effects found with animate musical sources are thus comparable to those obtained in the visual domain in Heider and Simmel's abstract animations, which produce the impression that geometric shapes are animate objects trying to achieve certain goals, as we saw in (13) above.

• Second, the music is usually understood to be itself, as whole, an intentional product. Both its form and the meanings it conveys can be attributed to an intentional agent. Let us call it the Narrator, in order to distinguish it from the 'real' composer, of which the listener might know nothing (this is of course the same distinction that one needs in literary theory between the writer and the narrator).

• Third, the music is normally interpreted by intentional agents, the musicians (computer-generated music might be perceived differently). And these may sometimes produce effects that are inconsistent with either of the first two levels.

As a point of comparison, think of a text with dialogues, read by a single actor. The three levels we have discussed will be present as well. First, the text is about some objects, some of them intentional. In particular, when the actor reads dialogues, he will successively incarnate the various characters. Second, the text is presented as being written by a narrator, who may come to the forefront if some passages are written in the first person. In a paragraph that does not involve dialogues, some of the interpretive choices made by the actor (involving for instance a happy or a sad tone of voice) might thus be assigned to the narrator. Finally, the actor's own intentionality is by force present as well. While a good actor might 'disappear' behind his character, this may fail to be the case, by design or by accident. If a prop malfunctions, the actor might say something *qua* actor; or he might express disapproval as an audience member makes disruptive noises.

The same distinctions arguably hold in music. To make things concrete, let us consider a group ending with a crescendo, which may often be interpreted as an intentional signal that a goal has been reached. If one artificially modifies a midi version of Mahler's Frère Jacques in such a way that we add a crescendo by the horn on the very last part of the last note of the piece (a D) [https://soundcloud.com/philippeschlenker/mahler-frere-jacques-last-4-final-cresc], we get an effect which is in remarkably bad taste, but is easily interpretable: the horn player is triumphantly indicating that the end of the piece has (finally) been reached. In this case, the voices all finish decrescendo, so that this final crescendo can't coherently be attributed to them. Nor is it natural to think that the narrator somehow 'wants' this final crescendo, which contradicts the musical intention that can be inferred from the decrescendo of the last bars (the procession is moving away, or at least its sound is gradually dying out). Thus one can only attribute this triumphant outburst to the musician – which also explains why the effect is in such bad taste.

8.3 Dialogues

Up to this point we have assumed that there exists one narrator per musical piece. But once music is endowed with a semantics, there also exists the possibility that a piece could involve a dialogue between different narrators. This possibility might in particular be instantiated in chamber music, with each instrument corresponding not to a voice but to a narrator. Detailed work would be needed to distinguish – possibly on a case by case basis – among two possible interpretations:
• One is that each instrument corresponds to a voice, hence produces the auditory trace of an object. This would still allow the voices to appear as intentional and to interact in complex ways (as a comparison, one may think of dancers that interact with each other but don't thereby *talk* to each other).
• An alternative is that each instrument corresponds to a narrator and that there is a genuinely a dialogue between them (here the point of comparison should be actors involved in a dialogue).
Thus the precise analysis of the various levels of intentionality involved in a piece, as well as the musical means by which they are realized as intentional, could prove illuminating in future studies of music semantics.

9 Emotions

The semantic content of music is often given in terms of emotions. These have so far been absent from our discussion. Can they find a natural place in our analysis, or should it be modified to accommodate them?

9.1 Emotional levels

Certainly our framework has a place for *some* emotional effects. In fact, just as was the case with attributions of intentionality, attribution of emotions may take place at several levels:
(i) the objects corresponding to the pseudo-sources of the music may be perceived as emotional (as a point of comparison, think of paintings that represent happy or sad individuals);
(ii) the narrator may be understood to express his own emotions; and in case a musical piece is interpreted as a dialogue among different narrators, each narrator may express his own emotions as well;
(iii) finally, the musician may express his own emotions, although this could generally be expected to be in bad taste if it goes against the music.

9.2 Means of expression

As in the other semantic effects, we have been discussing, there are two sources of inferences on emotions: properties of standard auditory cognition; and specifically tonal properties.

9.2.1 Inferences from standard auditory cognition

Let us start with some examples. All other things being equal, it would seem that greater happiness is attributed to a source which uses higher pitch and changes more quickly. Simple manipulations of Mahler's Frère Jacques display the effect: the effect of a funeral procession is lost as the music is raised in pitch and in speed, as seen in (69). We conjecture that similar effects would be obtained with non-musical voice, for instance, with greater speed and higher pitch (for a given voice) associated with greater animation and possibly happiness.

(69) Mahler's Frère Jacques, measures 3-6
a. Original version
https://soundcloud.com/philippeschlenker/mahler1-3ext-orchestra-beginning-normal

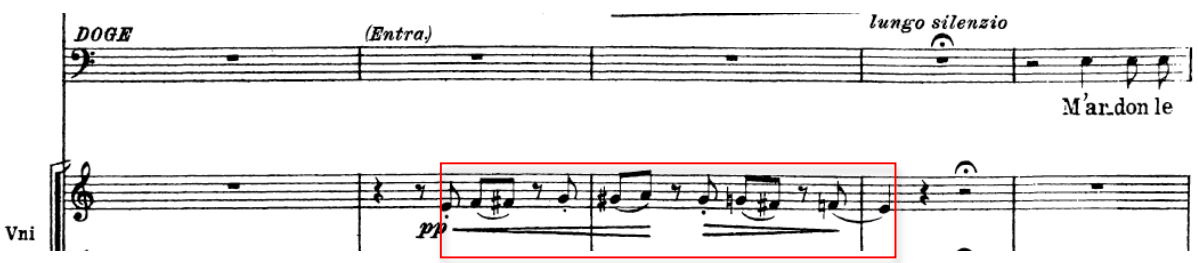
b. Original version, 2.5 times as fast
https://soundcloud.com/philippeschlenker/mahler1-3ext-orchestra-beginning-speed25
The impression of solemnity disppears
c. Original version, 2 octaves up
https://soundcloud.com/philippeschlenker/mahler1-3ext-orchestra-beginning-2-octaves-up

d. Original version, 2 octaves up, 2.5 times as fast
https://soundcloud.com/philippeschlenker/mahler1-3ext-orchestra-beginning-2-octaves-up-speed25

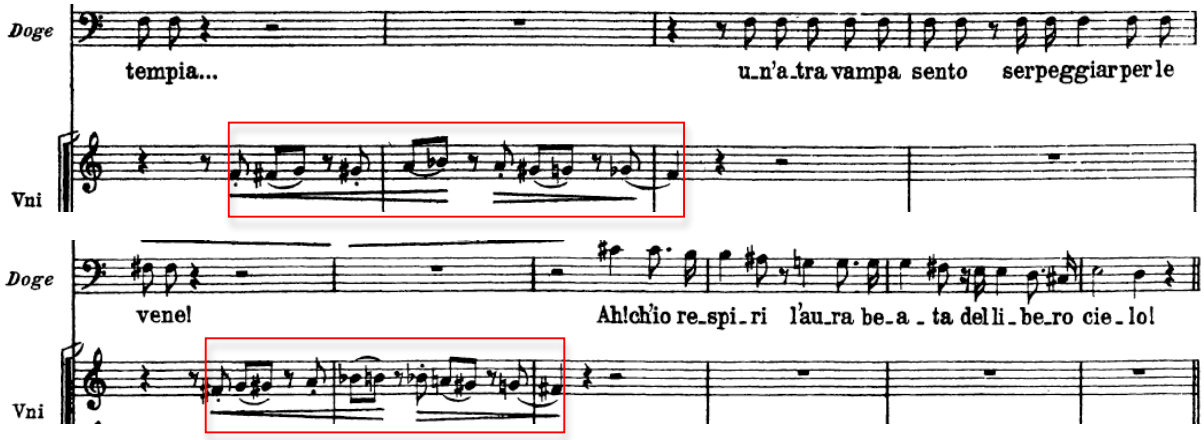
The piece seems much happier than in the original version

As we discussed in connection with Strauss's *Zarathustra*, powerful semantic effects can be produced by upwards or downwards melodic movement alone (en when a piece is entirely re-written in C's). An emotional example is provided by the end (Act III, Scene 3) of Verdi's *Simon Boccanegra*: three chromatic cycles evoke rising and receding effects of the poison that Simon drank in Act II. Each of the boxed sequences in (70) is made of 2 ascending chromatic sequences[29] in eighth notes (e.g. E F F#; G G# A), followed by one descending sequence with a similar rhythm (e.g. G# G F#), and a 2-note sequence (e.g. F E) ending on a longer note – the very same one that had started the cycle. The following cycles follow the same pattern, raised each time by a half-tone. The effect produced is arguably to evoke three cycles of Simon's increasing discomfort, by way of a mapping between the musical source and the intensity of Simon's discomfort.

(70) Verdi - Simon Boccanegra, Act III, Scene 3: (partial score: Simon and violins)
https://youtu.be/8F9Otx-wee8
'My head is burning, I feel a dreadful fire creeping through my veins...'



[29] In the full score, the melody is doubled in fourths and sixths.



In a non-musical domain, Aucouturier et al. 2016 showed that acoustic manipulations of a voice can significantly affect the emotions it conveys. The effects are illustrated in (71). One of the means to effect one of the manipulations, involving the 'afraid' condition, involves a vocal version of vibrato, as seen in (71)d.[30]

(71) Effect of voice manipulation on the perception of emotions (from Aucouturier et al. 2016)
a. Natural male voice [French]
http://www.pnas.org/content/suppl/2016/01/05/1506552113.DCSupplemental/pnas.1506552113.sa01.wav

b. 'Happy' manipulation:
it "modifies its dynamic range using compression to make it sound more confident and its spectral content using high-pass filtering to make it sound more aroused"
http://www.pnas.org/content/suppl/2016/01/05/1506552113.DCSupplemental/pnas.1506552113.sa02.wav

c. 'Sad' manipulation:
it 'operates on pitch using downshifting and spectral energy using a low-pass filter and a formant shifter'
http://www.pnas.org/content/suppl/2016/01/05/1506552113.DCSupplemental/pnas.1506552113.sa03.wav

d. 'Afraid' manipulation:
it 'operates on pitch using both vibrato and inflection'[31]
http://www.pnas.org/content/suppl/2016/01/05/1506552113.DCSupplemental/pnas.1506552113.sa04.wav

Now it is likely that whatever explains these musical effects with voices will trigger related effects. While it is certainly not the case that vibrato in music always produces an impression of fear, it does seem to be the case that vibrato is associated with heightened emotions – possibly because it is suggestive of decreased control by the source. Be that as it may, it seems likely that the emotional effect produced by vibrato is at least in part derived from effects that arise in non-musical sounds such as human voice.

Juslin and Laukka 2003 propose a general theory in which "music performers are able to communicate basic emotions to listeners by using a nonverbal code that derives from vocal expression of emotion". In a review of multiple studies, they argue that similar cues are used in the vocal and in the musical domain to express a variety of emotions, as summarized in (72). The parallelism between the vocal and the musical domain is expected from the perspective of a source-based semantics in which inferences about the emotional state of a source (or for that matter of a musical narrator) are drawn in part on the basis of normal auditory cognition. In addition, Sievers et al. 2013 suggest that there are homologies between the mechanisms that trigger emotions in the musical and in the visual domain, which should put interesting constraints on a music semantics.

[30] Spectacularly, Aucouturier et al. show that when subjects hear their own voice through earphones, if their voice is manipulated in real time through these means, subjects mostly fail to detect something abnormal, but the manipulation *affects their own emotional state* – as if they monitored it by voice clues.
[31] "Vibrato was sinusoidal with a depth of 15 cents and frequency of 8.5 Hz. Inflection had an initial pitch shift of +120 cents and a duration of 150 ms." (p. 4)

(72) Juslin and Laukka 2003

Summary of Cross-Modal Patterns of Acoustic Cues for Discrete Emotions

Emotion	Acoustic cues (vocal expression/music performance)
Anger	Fast speech rate/tempo, high voice intensity/sound level, much voice intensity/sound level variability, much high-frequency energy, high F0/pitch level, much F0/pitch variability, rising F0/pitch contour, fast voice onsets/tone attacks, and microstructural irregularity
Fear	Fast speech rate/tempo, low voice intensity/sound level (except in panic fear), much voice intensity/sound level variability, little high-frequency energy, high F0/pitch level, little F0/pitch variability, rising F0/pitch contour, and a lot of microstructural irregularity
Happiness	Fast speech rate/tempo, medium–high voice intensity/sound level, medium high-frequency energy, high F0/pitch level, much F0/pitch variability, rising F0/pitch contour, fast voice onsets/tone attacks, and very little microstructural regularity
Sadness	Slow speech rate/tempo, low voice intensity/sound level, little voice intensity/sound level variability, little high-frequency energy, low F0/pitch level, little F0/pitch variability, falling F0/pitch contour, slow voice onsets/tone attacks, and microstructural irregularity
Tenderness	Slow speech rate/tempo, low voice intensity/sound level, little voice intensity/sound level variability, little high-frequency energy, low F0/pitch level, little F0/pitch variability, falling F0/pitch contours, slow voice onsets/tone attacks, and microstructural regularity

Note. F0 = fundamental frequency.

9.2.2 Inferences from tonal properties

But strong emotional effects are also produced by specifically tonal properties of music. As is well-known, the major version of a piece typically produces a happier impression than its minor counterpart, as can be seen in the major counterparts in (73) of the four (minor) realizations of the beginning of Mahler's Frère Jacques already discussed in (69).

(73) Mahler's Frère Jacques, measures 3-6 – major transposition

a. Original tempo and pitch
https://soundcloud.com/philippeschlenker/mahler1-3ext-orchestra-beginning-normal-major

b. Original version, 2.5 times as fast
https://soundcloud.com/philippeschlenker/mahler1-3ext-orchestra-beginning-speed25-major

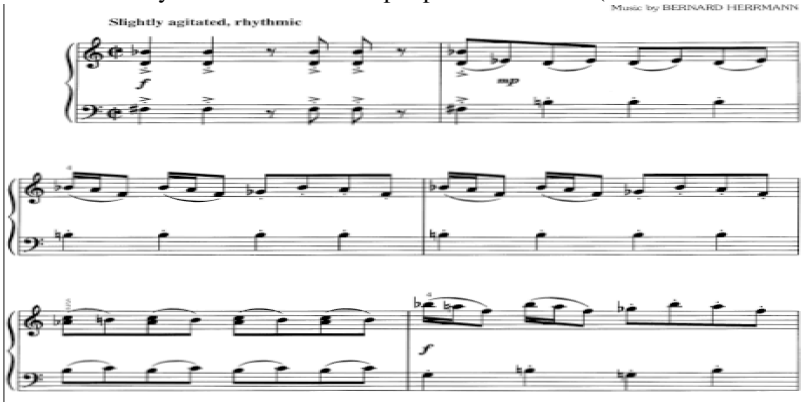
c. Original version, 2 octaves up
https://soundcloud.com/philippeschlenker/mahler1-3ext-orchestra-beginning-2-octaves-up-major

d. Original version, 2 octaves up, 2.5 times as fast
https://soundcloud.com/philippeschlenker/mahler1-3ext-orchestra-beginning-2-octaves-up-speed25-major

It is safe to assume that in each case the major version sounds happier and/or more assertive than its minor counterpart. Why is this so is debated, but it is likely that the fact that a major chord (e.g. C E G) is more consonant than a minor chord (e.g. C Eb G) plays a role in this.

We saw in our discussion of Saint Saëns's Tortoises (in (30)) that a dissonance can be interpreted as a point of physical disequilibrium. But it is often interpreted in music in terms of *emotional* disequilibrium. In fact, the chromatic progressions in (70) already had this character: the effect of rising discomfort largely disappears if the piece is re-written in diatonic rather than chromatic terms, as is done in Appendix II. But a far more extreme example is afforded by Herrmann's music for Hichcock's *Psycho*; a simplified piano reduction is given in (74). Strikingly, it starts with a D F# Bb (augmented fifth) chord, which sounds dissonant – and is preserved over the first half of the second bar. Of course various other choices contribute to the impression of mental imbalance, including the ostinato of the basic melodic movement, and the rhythm.

(74) Herrmann's Psycho – Prelude – simple piano reduction (Publisher: Hal Leonard)



Still, the dissonances play a crucial role in the effect obtained, as can be seen if the original version (in a more complete score) is compared with two modifications that eliminate the dissonances. Both are written in the 'closest' key to Herrmann's original, G minor. The original version is striking for the feeling of anguish that it produces; much is lost in the rewritten versions.

(75) Herrmann's Psycho - reduction in (74), re-written in G minor (A. Bonetto)[32]
a. Original redution
https://soundcloud.com/philippeschlenker/herrmann-psycho-base-1-15

b. Same as in a., re-written in G minor without dissonances)
https://soundcloud.com/philippeschlenker/herrmann-psycho-v1a-1-15

c. Same as a., closer to the original harmony
https://soundcloud.com/philippeschlenker/herrmann-psycho-v1b-1-15

9.3 Experienced events vs. objective events

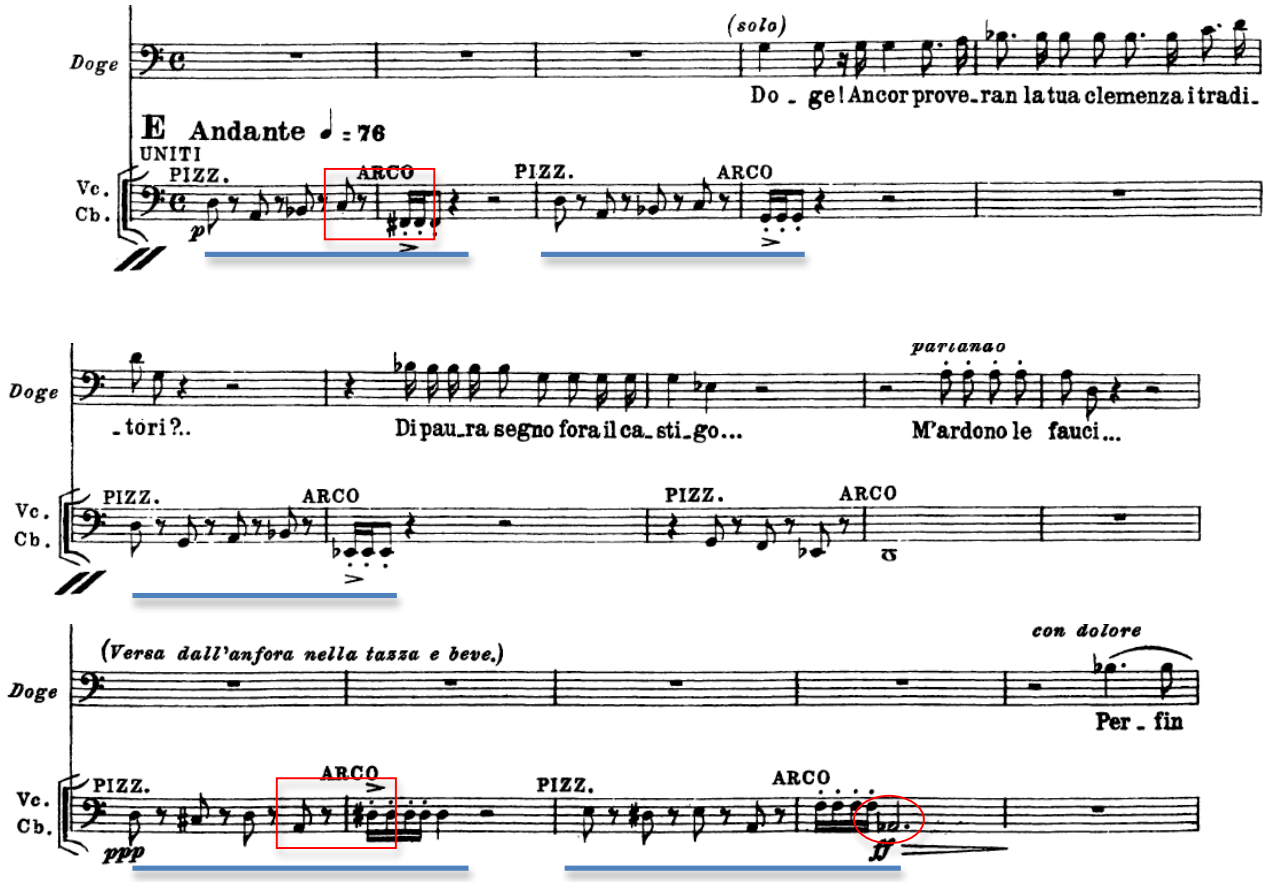
When one listens to Herrmann's *Psycho*, one does not just perceive the emotions of a source or of the musical narrator. Rather, one's own emotions seem to be affected. This is one sense in which music is often thought to bear a special relation to emotions. Now part of this effect can probably be analyzed as an instance of 'emotional contagion': one may *feel* sad when observing someone who looks sad. But there might be something more fundamental going on. Since our analysis leaves entirely open what the sources of the music are conceived to be, there is the option of treating them as *experienced* sources. In other words, it would make much sense to take the objects and events that our analysis posits to be *experienced* objects and events rather than objective ones. This would make it possible to associate with voices series of experienced events, which may be partly or entirely internal. The existence of the tactus probably favors such 'internal' interpretations of the music. Assuming that it is interpreted in terms of regular impulses of energy, it corresponds to a standard part of internal experience, involving for instance breathing, heartbeats, or just walking.

9.3.1 An example

An example from Verdi's *Simon Boccanegra* will make this general point concrete. In Act II, Scene 8, Simon drinks a cup which, unbeknownst to him, has been filled with poisoned water; consequences in Act III were discussed above in (70), when Simon begins to feel the effects of the poison. But even before he drinks from the cup, the cello theme makes clear that something momentous and disturbing is happening, as seen in (76). Crucially, the only character present, Simon himself, is unaware of what is going on, hence the music cannot serve to evoke his own emotions. Rather, it is probably the viewer's own emotions which are now reflected in the music (and possibly also the forces of destiny).

[32] In greater detail, the transformations were as follows:
(i) From (75)a to (75)b: **Bar 1**: F# > G **Bar 2**: F# > G ; B > Bb **Bars 3-4/6-7** : F > G ; Gb > G ; B > Bib **Bar 5:** C > D; B > Bb ; Ab > G ; Eb > D
(ii) From (75)a to (75)c: same as (i), but the boxed F > G in (i) becomes F > F# instead.

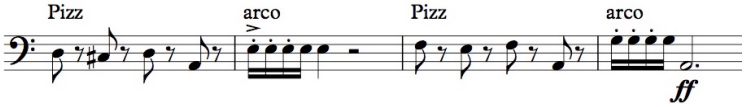
(76) Verdi's Simon Boccanegra, Act II, Scene 8:
https://youtu.be/LIHh1QrM34I



Several means conspire in the cello theme (underlined five times) to yield the impression that something momentous and disturbing is happening. The entire passage is in minor keys (arguably G minor in the first two lines and D minor in the last line). In addition, there in an alternation between slow eight notes, with pizzicato timber, and fast sixteenth notes, arco, played with an initial accent: this evokes ordinary and light events followed by faster and heavier events combined with an impulse of energy. In the two boxed passages, the interval separating the slow eighth notes from the fast sixteenth notes is a 'tritone' (diminished fifth), which is rather dissonant. And the last line involves a gradual chromatic ascent, D D# E F, indicative of the dramatic development. Rewriting the last line in D minor without chromatic excursions (as in (77)b) suppresses the tritone interval, and removes much of the feeling of tension and anguish. Last but not least, the last five notes would lead one to expect a series FFFF F, but the fortissimo conclusion on a low Ab (circled) instead of an F indicates that the expected course of events has been disrupted. (In the version of Giorgio Gallione [https://youtu.be/LIHh1QrM34I], Leo Nucci as Simon drinks from the cup at exactly that point.)

(77) a. The last line of (76) is written with a chromatic ascent and a tritone interval (boxed), yielding a feeling of tension and anguish
https://soundcloud.com/philippeschlenker/verdi-boccanegrabase-poison

b. Rewriting a. in D minor (without chromatic excursions) removes much of the feeling of tension and anguish (re-written by A. Bonetto)
https://soundcloud.com/philippeschlenker/verdi-boccanegrav1-poison-d-minor



9.3.2 Necessary refinements of our framework

In such cases, our general framework could be applied, but only if we take the basic elements of our ontology to be experienced rather than objective elements – experienced in particular by the listener. How can this provision be incorporated into the formal analysis we sketched above? If we go back to

our toy model in (36), we could for instance state the Harmonic stability condition in a slightly more sophisticated fashion. Considering a voice associated with an object O, we assumed that when a musical event M_i is less harmonically stable than M_k, O is in a less stable position in the event e_i denoted by M than in the event e_k denoted by M_k, as is stated in (78)b. We could now specify that in this case either O is in a less stable position in e_i than in e_k, or O's being in e_i causes a less stable emotion than O's being in e_k. Thus the modified Harmonic stability condition, stated in (78)b, is disjunctive, a property it shares with our old Loudness condition, seen in (78)a.

(78) a. Loudness (= (36)b)
If M_i is less loud than M_k, then either:
(i) O has less energy in e_i than in e_k; or
(ii) O is further from the perceiver in e_i than in e_k.

b. Harmonic stability – Original version (= (36)c)
If M_i is less harmonically stable than M_k, then O is in a less stable position in e_i than it is in e_k.

c. Harmonic stability – Modified version
If M_i is less harmonically stable than M_k, then either:
(i) O is in a less stable position in e_i than it is in e_k; or
(ii) O's being in e_i causes a less stable emotion in the perceiver than O's being in e_k.

Of course many further adjustments should be made to the other conditions as well, to ensure for instance that an accent (with a peak of loudness, as in (76) above) may be interpreted in emotional rather than just object terms.

Finally, let us note that we were forced to stipulate certain properties of stability of real world events in our initial examples illustrating Harmonic stability. While simple cases may be intuitive enough, one would need to develop an independent theory of the 'stability' of real world events. When we make provisions for the possibility that musical voices denote series of experienced events that may be associated with all kinds of emotions, it becomes clear that a proper music semantics presupposes an understanding of the structure of these emotions – a non-trivial requirement.

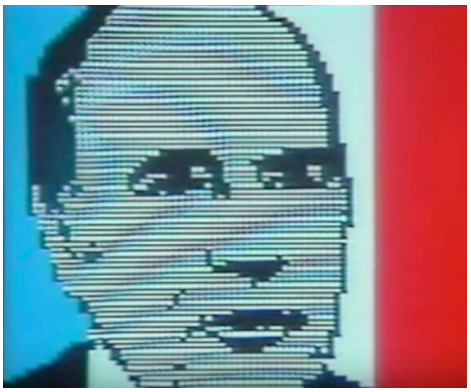
10 Extensions

We end this piece with further considerations about various potential extensions of our analysis.

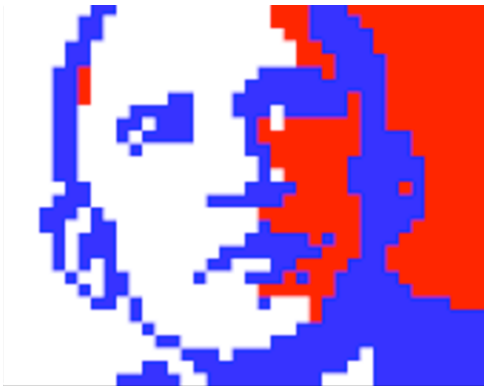
10.1 Context and granularity

In these *Prolegomena,* we only attempted to sketch the general form of a music semantics. One important issue in actual analyses will lie in determining the *level of granularity* of the interpretation. One may decide to take each and every musical event corresponds to a real world event. But often one may want to have a less fine-grained interpretation. The same issue arises when determining under what conditions a pictorial representation is compatible (i.e. could denote) a real world situation, as is illustrated with the coarse-grained pictures in (79).

(79) a. Coarse-grained pictorial representation of French President- François Mitterrand in 1981



b. Coarse-grained pictorial representation of US US President Barack Obama



Two mechanisms are crucial if we are to ensure that these pictures represent their intended denotations.
(i) First, we should make sure that the set of possible denotations is small enough – it may be restricted to the set of salient politicians in the situation.

(ii) Second, we should make sure that not all details of the pictures are required to correspond to something in the intended denotation. For instance, in a pixelized representation, some edges are due to the requirement that squares are used to represent shapes, and must be in part disregarded.

Both mechanisms should prove important in music semantics.

(i) First, semantic intuitions that would otherwise be very unclear can be sharpened by reducing the set of possible denotations. One way to do this is by way of titles or of explicit (linguistic) descriptions. This is an important device in 'program music'. We saw striking instances of this mechanism in Saint-Saëns's *Carnival*. For instance, the listener's interpretation of the dissonance in the piece on Tortoises is influenced by the assumption that the main voice denotes an animal and probably represents something that it does; access to the title of the movement makes the referential intention even clearer.

(ii) Second, when analyzing a piece one may decide to interpret each and every note as corresponding to a real world event, or one may adopt a more coarse-grained interpretation. For instance, if we were to ask of a given movement of a swan whether it makes true the beginning of Saint-Saëns's piece in (33), one may be satisfied if there is a series of two movements, the second of which leads into a new spatial area (thus interpreting the modulation). Or one may wish to find a much more precise correspondence between the piece – e.g. its precise melodic movement – and the scene it is purportedly true of. Of course a coarse-grained interpretation will make the piece true in many more situations than a fine-grained one.[33]

10.2 Interpreting a piece

If this analysis is on the right track, a musician interpreting a piece may make musical decisions that further specificy the semantic interpretation of the musical score. Ending a piece fortissimo may produce the impression that the source is intentional and is reaching a goal. Ending a piece diminuendo and rallentando may yield the impression that the source is gradually losing energy and dying out. Ending diminuendo without much altering the speed might yield the impression that the source is moving away. These are of course simplifications, but the general point is that the musical interpreter may and sometimes has make semantic decisions that the score might leave open. It will still be the case that there are a plurality of situations that the music is compatible with (true of), but the musician will usually reduce the set of situations that are compatible with a score.

10.3 Aesthetic considerations

We have been silent on aesthetic considerations – simply because it is one thing to set up a music semantics, and quite another to assess the aesthetic value of music. If successful, music semantics should come to explain why bad and good music alike produce semantic effects; it is not its goal to be a theory of music aesthetics. Still, one might hope that some aesthetic considerations might in the end build on insights gained from music semantics.

10.4 Semantic effects beyond music

The key idea of our semantic analysis of music is that denotational inferences may be drawn both from standard auditory cognition and from the tonal properties of music. This idea could in principle be applied to other areas as well.

First, one could ask whether a kind of visual counterpart of music could be devised. It would be based on animations that convey information by way of a combination of standard representational properties and ones that are internal to a more abstract system. A very simple example can be found in animated heat maps [https://www.youtube.com/watch?v=4_v2EZGiA7w]: while not quite direct, the geographical content of the map is based on standard principles of visual perception, modulo some simplifications (as a first approximation, a country is seen on a map as if it were perceived from very high up, say from space); simultaneously, there is a color code which is based on natural properties of colors: 'warm' colors (e.g. red) represent high degrees of the relevant property, and 'cool' colors low degrees. But of course the structure of colors is entirely different from the structure of tonal pitch space, and thus it is only at a conceptual level that a general correspondence between the auditory and the visual domain can be found.[34] (It would be interesting to explore in the future closer correspondence between the musical

[33] The issue of granularity should arise as well when music is explicitly set in correspondence with something else, as in singing in general and opera in particular, or dance accompanied with music. In such cases, one might strive for a very fine-grained correspondence, or be more liberal.

[34] Simple systems of music visualization give rise to abstract animations as well, but the transposition to a different modality only preserves some of the inferences triggered by the music. For instance, in its basic form, Stephen Malinowski's 'Music Animation Machine' [http://www.musanim.com/mam/overview.html] only encodes pitch and duration; loudness, for instance, is not represented. Coloration can be added to encode instrumentation or

and the visual domains, in particular by seeking to find visual counterparts of the various cues that trigger inferences about the sources, be they drawn from normal auditory cognition or from harmonic considerations.)

Second, one could ask whether related ideas could be applied to the analysis of abstract painting. Certainly some standard principles of visual perception are at play in abstract painting – which is the reason we usually don't just see shapes on a canvas, but (possibly very abstract) objects – something that already played an important role in Heider and Simmel's abstract animations. It remains to be seen whether certain non-natural properties of the paintings could be interpreted in a way that is comparable to the tonal properties of music.

Finally, one might attempt to apply related ideas to dance, for two reasons: like music, it triggers referential and emotional inferences on the basis of 'natural' and more abstract properties of perception; and in addition, it is often coordinated with music – and hence one might ask how the two mediums are coordinated: do they give rise to a single semantic representation? We refer the reader to Charnavel 2016 and Napoli and Kraus, to appear, for initial attempts.

11 Conclusions

11.1 Theoretical conclusions

If our proposal is on the right track, music has a semantics, but one which is closer to picture semantics than it is to linguistic semantics. We have treated music cognition as being continuous with normal auditory cognition, and in both cases we took the semantic content of an auditory percept to be closely connected with the set of inferences it licenses on its causal sources, analyzed in appropriately abstract ways (e.g. as 'voices' in some Western music). However music semantics is special in that it aggregates inferences from two main sources: normal auditory cognition, but also tonal properties of the music. This made it possible to sketch a truth-conditional semantics for music: a music piece *m* is true of a series of events (undergone by an object) just in case there is a certain structure-preserving map between the musical events and the real world events they are supposed to denote.

Several consequences of this semantic approach that should be explored in future research. First, aspects of musical syntax can arguably be reconstructed on semantic grounds. In particular, we argued that grouping structure can be seen to reflect the mereology of the denoted events, and we tentatively suggested that even the headed nature of Lerdahl's and Jackendoff's time-span reductions could be reinterpreted in semantic terms (we left for future research an exploration of the role that prolongational reductions should play in this debate). Second, we argued that our source-based framework is versatile enough to find a place for intentional effects at various levels, and we made a similar suggestion about emotional effects, arguing that the general framework might account for the special connection between music and emotions without necessarily requiring major additions.

11.2 Methodological conclusions

Finally, although our proposal owes little to linguistic *principles*, it relied on linguistic *methodology* in its attempt to construct minimal pairs to display semantic effects (this methodology is, more broadly, that of 'controlled experiments', be they based on 'real' experiments or on introspection). Once a potential semantic effect was identified, and a hypothesis formulated as to its origin, the analysis could be tested by isolating the crucial parameter, in one of two ways: in most cases, we discussed minimal musical pairs that differed whenever possible by just one parameter, the crucial one, in order to show that the target semantic effect was weakened when the parameter was distorted; in a couple of cases, we were able to abstract away (rather than control for) other parameters, as when we 'rewrote' a passage with C's only in order to display the specific effect of (non-harmonic) melodic movement (this method is more standard when rhythm is investigated, as it is easy to keep the rhythm of a passage while neutralizing the melody and harmony). For obvious reasons, the method of

---

harmony. Harmonic coloring [http://www.musanim.com/mam/circle.html] has been used to provide, for instance, an animated rendition of Stravinsky's *Rite of Spring* [https://www.youtube.com/watch?v=5IXMpUhuBMs]. It is immediate that even harmonic coloring only yields a crude (and not necessarily intuitive) encoding of the complex harmonic relations among notes and chords.

(i) Stephen Malinowski's 'harmonic coloring' for his Music Animation Machine

minimal pairs requires the help of composers, in order to ensure that the pairs created are really as minimal as possible given the constraints of the relevant musical idiom. While we focused on the very simplest cases, particularly interesting issues will arise when the *interaction* between several cues is investigated, as it is in this more complex case that a source-based semantics will make it possible to make non-trivial predictions.

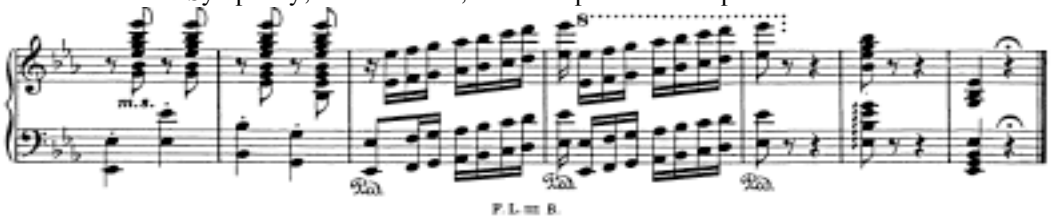
In order to *explain* semantic effects, methods differed depending on whether they had their origin in normal auditory cognition or in properties of tonal pitch space. In the first case, similar effects must be displayed in non-musical audition (and more broadly in perception). In the second case, explanations have to be more theory-internal, building on what one takes to be relevant properties of tonal pitch space. Importantly, the inferences that one might need to test are quite abstract in nature, hence in future studies great care should be devoted to the precise formulation of the inferential questions, and further methods should be developed to sharpen semantic intuitions – for instance by providing additional information (by way of titles, stories, or other non-musical information) so as to make inferences more precise.

Last, but not least, these preliminary investigations have been quite parochial, since they were restricted to a few pieces of Western classical music. A cross-linguistic investigation of music semantics should prove illuminating.

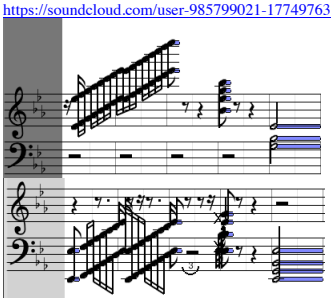
Appendix I. Finishing Downwards in Beethoven's Third Symphony

Beethoven's Third Symphony, here in Liszt's piano transcription, ends with a chord downwards, as shown in (80)a. One can artificially raise the final chord by two octaves, so that the piece resolves in the same way harmonically, but upwards, as in (80)b; the effect is rather less conclusive. If the final chord is raised by three octaves, as in (80)c, the effect becomes inconclusive.

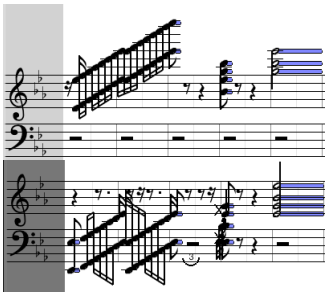
(80) Beethoven's Third Symphony, last measures, in Liszt's piano transcription



a. The original version ends forte but is conclusive, with two I chords in the last two measures, and a clear movement downwards.
https://soundcloud.com/user-985799021-177497631/beethoven-3rd-symphony-liszt-508-end-normal/s-wQpR1



b. If the last chord is raised by 2 octaves, the ending sounds less conclusive.
https://soundcloud.com/user-985799021-177497631/beethoven-3rd-symphony-liszt-508-end-last-2-oct-up/s-qbrd8



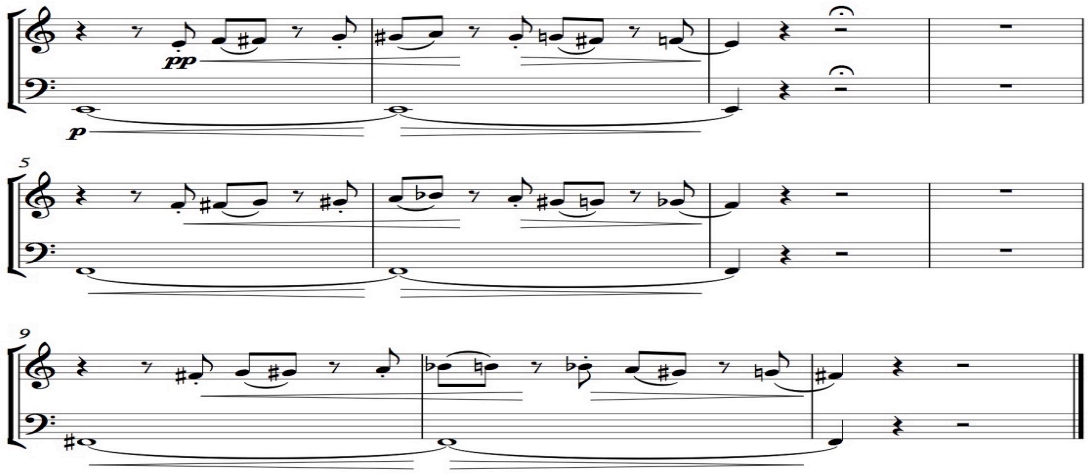
c. If the last chord is raised by 3 octaves, the ending sounds inconclusive.
https://soundcloud.com/user-985799021-177497631/beethoven-3rd-symphony-liszt-508-end-last-3-oct-up/s-l71Iv

Chromatic vs. diatonic progressions in Simon Boccanegra's poison scene

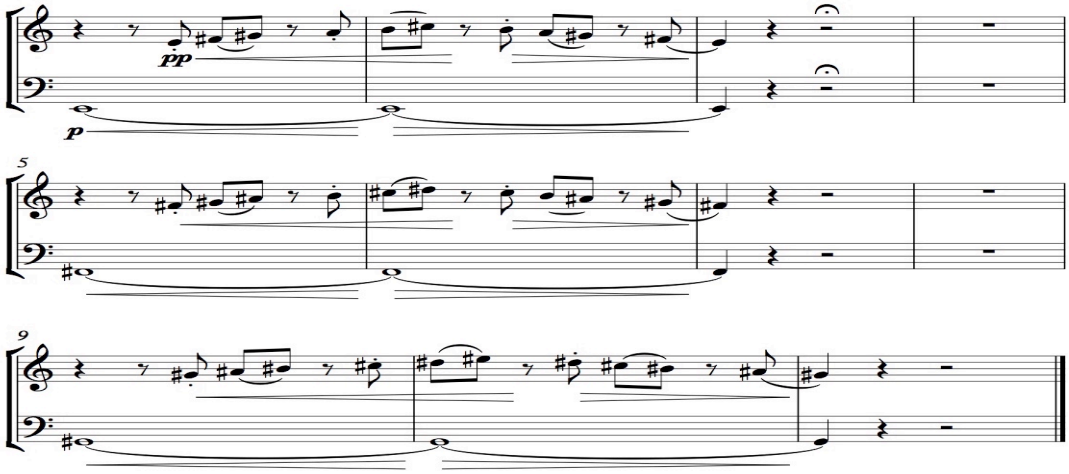
Besides pitch movement, harmonic considerations are crucial to the effects discussed in (70): the major version of the chromatic progression in (81)a fully loses the impression of rising discomfort of the original. The minor version only keeps it in part.

(81) Three versions of Simon Boccanegra's poison scene (b. and c. re-written by A. Bonetto)

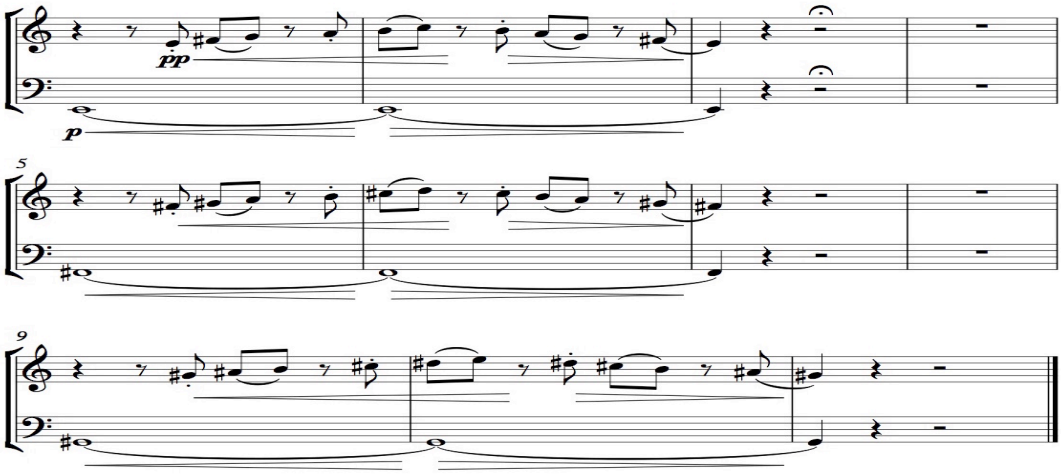
a. Simplified version of Simon Boccanegra's poison scene in (70)
https://soundcloud.com/philippeschlenker/verdi-boccanegra-poison-effect-base



b. Re-written version of a. in E major
https://soundcloud.com/philippeschlenker/verdi-boccanegra-poison-effect-major



c. Re-written version of a. in E minor
https://soundcloud.com/philippeschlenker/verdi-boccanegra-poison-effect-minor



References

Artstein, Ron: 2004 Focus below the word level. *Natural Language Semantics* 12(1): 1-22.

Aucouturier, J. J., Johansson, P., Hall, L., Segnini, R., Mercadié, L., & Watanabe, K. (2016). Covert digital manipulation of vocal emotion alter speakers' emotional states in a congruent direction. *Proceedings of the National Academy of Sciences*113(4):948-53. doi: 10.1073/pnas.1506552113.

Bregman, Albert S.: 1994, *Auditory Scene Analysis*. MIT Press.

Charnavel, Isabelle: 2016, First Steps towards a Generative Theory of Dance Cognition: Grouping Structures in Dance Perception. Manuscript, Harvard University.

Cross, I. and Woodruff, G. E.: 2008, Music as a communicative medium. In Botha, R. and Knight, C. (Eds.), *The Prehistory of Language*, Vol. 1, pp. 113–144.

de Vries, Mark: 2013, Multidominance and locality, *Lingua* 134(0), 149–169.

Desain, P., and Honing, H.: 1996, Physical motion as a metaphor for timing in music: the final ritard. In *Proceedings of the International Computer Music Conference* (pp. 458-460). International Computer Association.

Eitan, Zohar, and Roni Y. Granot: 2006, How music moves. *Music Perception* 23, 3:221-247,

Fitch, Tecumseh W., Reby, D.: 2001, The descended larynx is not uniquely human. *Proceedings of the Royal Society of London*. Series B, 268, 1669-1675.

Forte, Allen: 1959, Schenker's conception of musical structure. Journal of Music Theory. 3:1-30.

Greenberg, Gabriel: 2013. Beyond Resemblance. *Philosophical Review* 122:2, 2013

Halle, John: 2015, From Linguistics to Musicology. Notes on Structuralism, Musical Generativism, Cognitive Science, and Philosophy. In Brandt and Carmo (eds), *Music and Meaning, Annals of Semiotics 6/2015,* Presses Universitaires de Liège.

Huron, David: 2015. Cues and Signals: an Ethological Approach to Music-Related Emotion. In Brandt and Carmo (eds), *Music and Meaning, Annals of Semiotics 6/2015,* Presses Universitaires de Liège.

Jackendoff, Ray: 2009, Parallels and nonparallels between language and music. *Music Perception*, 26(3), 195-204.

Juslin P, Laukka P.: 2003, Communication of emotions in vocal expression and music performance: Different channels, same code? *Psychological Bulletin*; 129(5):770–814.

Koelsch S.: 2011, Towards a neural basis of processing musical semantics. *Physics of Life Reviews* 8(2):89–105

Kominsky, J. F., Strickland, B., & Keil, F. C. *Sensitivity to Newtonian regularities in causal perception: Evidence from attention*. Poster presentation, Vision Sciences Society annual meeting, May 16-21, 2014.

Larson, Steve: 2012, *Musical Forces: Motion, Metaphor, and Meaning in Music*. Indiana University Press.

Lemasson Alban, Ouattara Karim, Bouchet Hélène and Zuberbühler Klaus, 2010. Speed of call delivery is related to context and caller identity in Campbell's monkey males. *Naturwissenschaften* 97 (11): 1023-1027.

Lerdahl, Fred and Ray Jackendoff: 1983, *A generative theory of tonal music*. Cambridge, MA: MIT Press.

Lerdahl, Fred: 2001. *Tonal Pitch Space*. Oxford University Press.

McCawley, James D. (1982): 'Parentheticals and Discontinuous Constituent Structure', *Linguistic Inquiry*13(1), 91–106.

Meyer, L.B.: 1956, *Emotion and Meaning in Music*. University of Chicago Press, Chicago

Napoli, Donna Jo and Kraus, Lisa: to appear, Suggestions for a parametric typology of dance. *Leonardo*. doi:10.1162/LEON_a_01079

Ohala, J. J.: 1994, The frequency code underlies the sound-symbolic use of voice pitch. In L. Hinton, J. Nichols & J. J. Ohala (Eds.), Sound Symbolism, 325- 347. Cambridge: Cambridge University Press.

Pesetsky, David and Katz, Jonah. 2009. The Identity Thesis for Music and Language. Manuscript, MIT.

Rooth, Mats. 1996. Focus. In *Handbook of Contemporary Semantic Theory*, ed. by Lappin Shalom, 271–297. Blackwell, Oxford.

Schwarzschild, Roger. 1999. GIVENness, AvoidF and other Constraints on the Placement of Accent. *Natural Language Semantics* 7(2): 141–177.

Sievers, B., Polansky, L., Casey, M., & Wheatley, T.: 2013, Music and movement share a dynamic structure that supports universal expressions of emotion. Proceedings of the National Academy of Sciences, 110, 70-75. doi:10.1073/pnas.1209023110

Varzi, Achille, "Mereology", The Stanford Encyclopedia of Philosophy (Winter 2015 Edition), Edward N. Zalta (ed.), URL = <http://plato.stanford.edu/archives/win2015/entries/mereology/>.

Wolff, Francis: *Pourquoi la musique?* Fayard 2015

Zacks, Jeffrey M., Tversky, Barbara, and Iyer, Gowri: 2001, Perceiving, remembering, and communicating structure in events. *Journal of Experimental Psychology: Genera*l, 130, 29–58.