Prolegomena to Music Semantics[*]

Philippe Schlenker[**]

Typos corrected, May 11, 2018

To appear in *Review of Philosophy & Psychology*.

Music consultant: Arthur Bonetto

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 derived from 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') might be derivable on semantic grounds from an event mereology ('partology'), which also explains some cases in which tree structures are inadequate for music (overlap, ellipsis). (vi) Intentions and emotions may be attributed at several levels (the source, the musical narrator, the musician), and we speculate on possible explanations of the special relation between music and emotions.

Keywords: music, music semantics, musicology, music syntax

Note: See fn. 1 for the use of audiovisual examples.

---

1   Introduction[1]

1.1   Goals

While the *syntax* of music has been studied in formal detail (e.g. Lehrdahl and Jackendoff 1983, Lerdahl 2001, Pesetsky and Katz 2009, and Rohrmeier 2011 for classical music, Granroth-Wilding and Steedeman 2014 for jazz), 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 of music', we mean *a rule-governed way in which music can provide information (i.e. license inferences) about some music-external reality*, no matter how abstract (see for instance Lewis 1970, Larson 1995, Heim and Kratzer 1998, and Schlenker 2010 for the linguistic case).[2] The 'no semantics' position might well be the Null Hypothesis: there is little initial reason to think that music has systematic representational capabilities, let alone denotations or truth conditions. 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 pertain to emotions that a piece may evoke.[3]

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 initial guiding intuition is that *the informational content derived from a musical piece is given by the inferences one can draw about its virtual sources*[4]. In salient cases, these virtual sources are associated with the 'voices' of classical music theory: these voices structure the musical form, and the virtual sources we posit behind them serve as their denotations and provide their semantic content. This guiding intuition will have to be refined, however, because some of the informational content of the music is due to the movement of the virtual sources in tonal pitch space. Our analysis is thus developed 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 sound semantics* (we will sometimes say that the musical surface is the 'auditory trace' of some external events). In this respect, we initially treat musical 'signs' as Peircian 'indices' because their semantics is derived from a causal relation between sounds and their sources.[5] But this is only a first approximation, for 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.

---

[1] Audiovisual examples have been included in the text, numbered as **AV00**, **AV01**, etc. They can be accessed by clicking on hyperlinks, or alternatively by downloading the entire folder of audiovisual examples: http://bit.ly/2DBhNH6. In case technical problems arise, a manuscript with hyperlinks can be downloaded at https://ling.auf.net/lingbuzz/002925.

[2] This notion of semantics corresponds to what Koelsch 2012 calls 'extra-musical meaning'.

[3] For a recent critical discussion of some 'no semantics' views, see Larsen 2017 (p. 28), who cites (and seeks to refute) the following opinion by Kivy 1990: "in the long run syntax without semantics must completely defeat linguistic interpretation. And although musical meaning may exist as a theory, it does not exist as a reality of listening".

[4] The term 'virtual source' is due to Bregman, e.g. Bregman 1994. See also Nudds 2007 for an analysis of auditory cognition in terms of source perception.

[5] Peirce's tripartition includes icons, indices and symbols. Indices are representations "whose relation to their objects consists in a correspondence in fact" (Atkin 2010, Peirce 1868). By contrast, icons involve a 'likeness' between the representations and their objects. Whether musical signs in our analysis should also be treated as icons depends on how 'likeness' is defined; we come back to this point in Section 7.2. See also Koelsch 2011, 2012 for a discussion of Peirce's tripartition in a musical context.

Second, we take further inferences to be drawn from the behavior of the virtual sources with respect to tonal pitch space.[6] This space has non-standard properties (which differ across cultures), 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 with respect to that space.

The main challenge in what follows will be to prove the existence of these two types of musical inferences (inferences from normal auditory cognition, and tonal inferences), and to sketch a formal framework to aggregate them.

1.2   Theoretical directions

We take the present analysis to integrate two intuitions that were developed in earlier theories.

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' in 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 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 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 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 gives 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.

To see a very simple example, both kinds of inferences can be used to signal the end of a piece. One common way to signal the end is to gradually decrease the loudness and/or the speed. While this device could be taken to be conventional, it is plausible that it is in fact derived from normal auditory cognition: a source that produces softer and softer sounds, and/or produces them more and more slowly, may be losing energy.[7] But on the tonal side, it is also standard to mark the end of a piece by a sequence of chords that gradually reach maximal repose, ending on a tonic. Plausibly, an inference is drawn to the effect that a virtual source that manifests itself by a tonic is in the most stable physical position, with no tendency to move any further. Thus these two types of inference combined conspire to signal the end of a piece.

---

[6] With the addition of tonal inferences, we move further away from an analysis in terms of mere Peircian indices.

[7] While the notion of 'energy' should be further explicated, we can rely at this point on an intuitive notion of folk psychology, according to which objects are taken to have different levels of energy depending on their movements and more generally on their behavior.

It will be essential to develop an appropriately abstract analysis, for even when inferences are drawn on the basis of normal auditory cognition, they need not come with the requirement that the virtual sources are sound-producing: music can be used to evoke silent events, such as a sunrise; and the inference that the virtual source is gradually losing energy need not come with the requirement that the source is sound-producing. While we will informally develop our analysis in inferential terms, by collecting appropriately abstract inferences triggered by a musical piece (some of them based on normal auditory cognition, others on properties of tonal pitch space), we will also 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 *e* of events 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 cases, 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).[8]

1.3   Organization

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 no semantics. It thus takes a detailed empirical argument to show that a semantic approach is legitimate. In Section 3, we provide an initial example 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'. Having developed the core of our semantic analysis, we pause in Section 7 to reflect on its relation to logical semantics and to iconic semantics. We then consider extensions of the analysis. We argue in Section 8 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 9, we explore various levels of pragmatic analysis in music, before speculating in Section 10 on the role of musical emotions, and drawing some conclusions in Section 11. (Further technical details, extensions and speculations are found in four Appendices.)

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 (i.e. a deflationary analysis) 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

---

[8] Our analysis builds on further insights that have been developed in the literature on music cognition, and which might also find a place in the present framework. One influential line of inquiry takes various semantic inferences to be based on the attribution of animacy and intentions to some musical elements such as pitches, chords, and motives (Lerdahl 2001, Maus 1988, Monahan 2013). A second but related line takes important semantic inferences to be triggered by sound properties found in animal signals and/or in human speech (Cook 2007, Cross and Woodruff 2008, Blumstein et al. 2012, Bowling et al. 2010, Huron 2015, Ilie and Thompson 2006, and Juslin and Laukka 2003). Both directions are compatible with Bregman's general enterprise, and our source-based semantics makes important use of their insights, but in the general case it does not require that the virtual sources should be animate. A third line of investigation takes music to trigger inferences about movement (Clarke 2001, Eitan and Granot 2006, Larson 2012, Saslaw 1996) – which is compatible with the analysis of musical meaning as a 'journey through tonal pitch space'. Our source-based semantics allows the virtual sources to move in space, but it allows for many other types of events as well. Relatedly, a fourth line of investigation takes certain properties of music – e.g. the 'final ritard' that signals the end of a piece – to imitate properties of forces and friction in the natural world (Desain and Honing 1996, Honing 2003, Larson 2012). Within our framework, these are particular ways of triggering semantic inferences, but there are many others as well.

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 specific 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.

2.1 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 position 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, and Rohrmeier 2011, Pesetsky and Katz 2009, and Granroth-Wilding and Steedman 2014 have further contributed to this topic.

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 formalism 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 → L, L S
       L → la lu, la li
       (ii) Regular grammar:
       (la lu ∪ la li)*

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

This very simple language will serve as a useful point of comparison for later discussions: although these sequences of syllables do not have a semantics in the usual sense, they convey information (about their own form), hence some expectations and some pragmatic effects. In addition, one can *endow* this language with a pseudo-semantics pertaining to its form, which will serve as a useful point of comparison for some 'internal semantics' proposed for music (this point is revisited in Appendix I).

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* to anything extra-musical. A slightly different view is that music has a semantics, but one that 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 analysis 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, Wolff 2015 and Koelsch 2012 for a broader discussion of this general debate). We will argue that music has a semantics in the usual sense: it conveys information about a music-external reality. Thus we will not further discuss the 'no semantics' or the 'internal semantics' view in the main part of this article. But since the 'internal semantics' view has important proponents in music cognition, we revisit it in Appendix I.

2.3 Expectations and pragmatics

Even if music has no semantics (not even an 'internal' semantics), it certainly leads to all sorts of expectations, namely about its form; these expectations could be taken to constitute the 'meaning' of

music. Furthermore, music certainly conveys some type of information, namely about its own form; this suggests that music could in principle make use of certain devices to structure this information in optimal ways – one aspect of 'music pragmatics'. The meaning of music is often taken to lie in or even to be exhausted by such internal informational effects, and it is thus important to distinguish them from genuine semantic phenomena.

Meyer 1956 argued that "one musical event (...) has meaning because it points to and makes us expect another musical event" (Meyer 1956, chapter I). The resulting expectations and emotions lead to what Meyer calls 'embodied meaning'. For his part, Huron 2006 argues that various emotions of a musical or extra-musical nature derive from general properties of expectation, i.e. of our attempts to anticipate what will come next, in music or elsewhere. For Huron, "the emotions evoked by expectation involve five functionally distinct physiological systems: imagination, tension, prediction, reaction, and appraisal" (p. 7); he thus seeks to derive musical emotions from the interaction of these systems with musical anticipations (the resulting theory is called 'ITPRA', which is the acronym of the five physiological systems). Importantly for our purposes, Huron's analysis need not depend on the existence of a music semantics, which pertains to the relation between music and a music-external reality.

Certainly musical expectations have numerous effects on the listener, but they should not be confused with a semantics as we use the term here: these expectations by themselves do not allow music to convey information about a music-external reality. To go back to our linguistic analogy involving the meaningless syllables *la lu la li,* the syntax we defined in (1) leads to some expectations, for instance that the syllable *la* should be followed by *li* or *lu*, and that *li* as well as *lu* should be followed by *la*. Whether or not these expectations lead to emotions on the perceiver's part, they are entirely different from a *bona fide* semantics.

Similarly, the composer or performer may choose to highlight some aspects of the music, thus helping the listener to structure musical information in the intended way – which is one aspect of 'music pragmatics'. But this does not suffice to yield a semantics. To have a linguistic point of comparison, consider how language makes new elements salient (e.g. Schwarzschild 1999). In (3)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 of greater loudness, higher pitch and longer duration). If another element is focused instead, as is the case in (3)b, the result is deviant (as indicated by the sign #).
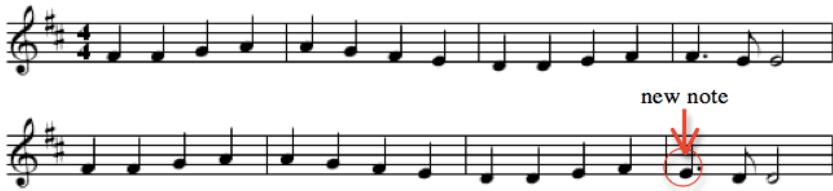
(3)   a. He will introduce me to her, and then he will introduce YOU to her.
      b. #He will introduce me to her, and then he will introduce you to HER.

While in this case the *meaning* of the focused elements might be crucial, there are further cases in which *form alone* plays 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 (4) 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.

(4)   [la lu] [la lu] [la LI] [la lu]

Do we find such effects in music? We might, as is illustrated in (5), where we feel that a performer might want to add greater emphasis on the first new note (circled) of the consequent, possibly realized by greater loudness and longer duration. Importantly, there might be other reasons why such an emphasis is found – including 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 9.1).

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

2.4   Summary and outlook

In this section, we have made the following points:
(i) It is uncontroversial that there is a musical syntax.
(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) Musical form is constrained in ways that give rise to expectations, but these are entirely different from a *bona fide* semantics. Similarly, it is plausible that music has an information-theoretic pragmatics, in the sense that one may highlight some aspects of musical form. But this does not entail that music has a semantics in the usual sense: even meaningless strings of syllables are naturally produced with means, such as contrastive focus, which highlight aspects of their structure.

Since the Null Hypothesis according to which music has no semantics is plausible, we will need to give serious empirical arguments to justify the project of a music semantics. We will do so in two steps: first, we will suggest that inferential properties of ordinary sounds play a role in music; second, we will argue that further inferences are produced when we take into consideration specifically tonal properties.

We discuss examples that make the general program plausible in Section 3. Inferences derived from normal auditory cognition are analyzed in Section 4, and inferences drawn from tonal properties are then discussed in Section 5. While our arguments are entirely based on introspective judgments, we will allude to relevant experimental results in the course of the discussion, and we will outline at the end of each section the methods that could be used to establish experimentally the correlations we discuss.

3   Examples of semantic effects

3.1   A visual example

Since we wish to argue that a non-natural 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 (1944) 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 (video examples: http://bit.ly/2CR5AB2). Lerdahl's suggestion is 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".[10] In the visual domain, then,

---

[10] In Heider and Simmel's animations, the interpretation involves attributions of agency and intentions – for instance a triangle may appear to have a destructive behavior. But further and more basic properties can be attributed to abstract shapes as well. As an example, Kominsky et al. 2017 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 a reason: a ratio of 3/1 is consistent with causal laws of elastic collision, whereas a ratio of 1/3 is not (an example in the 'violation' condition can be seen in AV00
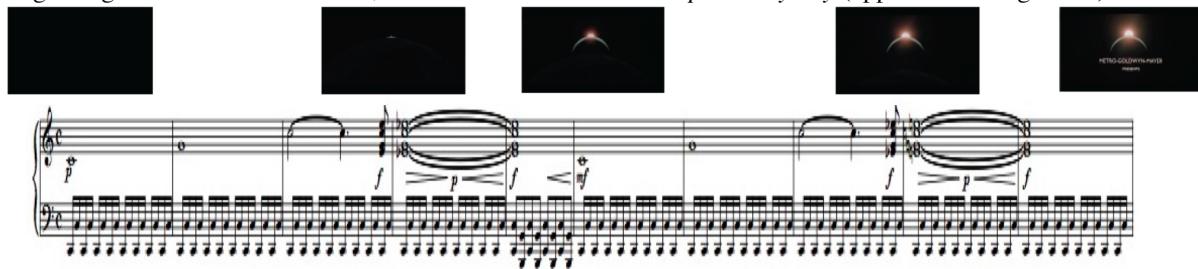
very abstract shapes can still give rise to inferences about virtual events that they are the 'visual traces' of.

3.2   A musical example

Let us turn to music, where sounds will play the role of 'auditory traces' of virtual events. Since the Null Hypothesis 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 they pertain to the development of phenomena in the extra-musical world.
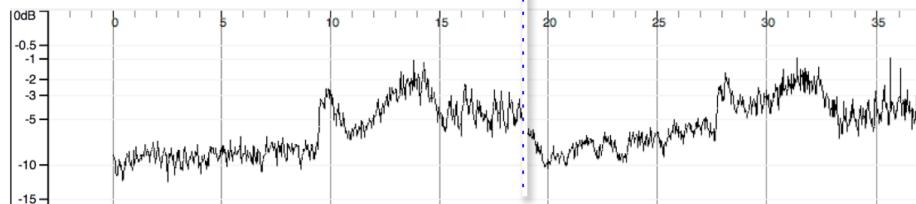
We consider the beginning of Strauss's Also Sprach Zarathustra ('Sunrise') [AV01 http://bit.ly/2FH39Ps], which is used as the sound track of the opening of the movie *2001: a Space Odyssey* [AV02 http://bit.ly/2DfiE3m]. In (6), we have superimposed some of the key images of the movie with a 'bare bones' commercial piano reduction (by William Wallace[11]). The correspondence already gives a hint as to the inferences one can draw from the music.

(6)   Beginning of Strauss's Zarathustra, with the visuals of *2001: a Space Odyssey* (approximate alignment)

Specifically, the film synchronizes with the music the appearance of a sun behind a planet, in stages – two of which are represented here. Bars 1-5 correspond to the appearance of the first third of the sun, bars 5-8 to the appearance of the second third (4-5 more measures are needed to complete the process – we simplify the discussion by focusing on the beginning). Now the music certainly evokes the development of a phenomenon in stages as well – which is unsurprising as it is (broadly speaking) an antecedent-consequent structure. But the music triggers more subtle inferences as well. A listener might get the impression that there is a 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 (7). In (7)a, we use chord notation to represent the harmonic development (with IM for a major I and Im for a minor I). In (7)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). Finally, in (7)c we use standard dynamics notation to encode loudness, using the dynamics (for the melody) in a richer piano reduction by Karl Schmalz[12].

(7)   a. Harmony: I   IM Im   I   Im IM
        b. Melody (soprano) 1   2   3   5 4   1   2   3   4 5

        c. Loudness: p   f>> p< f< mf   f>> p<   f<

---

http://bit.ly/2myIJAp). In this case, subjects seem to take the dots to be indicative of events that obey certain physical laws of the external world.

[11] Retrieved online on January 7, 2018 at http://www.8notes.com/scores/7213.asp. Dynamics were re-established by A. Bonetto on the basis of the orchestral score.

[12] Score retrieved online on January 8, 2018 at http://imslp.org/wiki/Also_sprach_Zarathustra,_Op.30_(Strauss,_Richard).

Harmonically, both the antecedent and the consequent display a movement from the first to the fifth to the first degree, 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 of a retreat at the end of the antecedent, as it reaches a stable position (I Major) and immediately moves 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 gesture features a crescendo, which produces the impression of a gradual development. Finally, each gesture 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. But even at this point, it is worth asking whether harmonic and melodic movement are *both* crucial to the observed semantic effect, in particular to the impression 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.

One way to test (i) [= same harmony without the melodic movement] 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 (6), further simplified to (8)a. In (8)b, two E's responsible for the melodic movement were removed (they are highlighted by arrows in (8)a). 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 because 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 seen in (8)c. While the effects are subtle, the comparison between these 'minimal pairs' suggests that although harmony plays an important role in the semantic effect we observe, the melodic movement might play a role as well.

(8)   a. A 'bare bones' piano reduction of the beginning of Strauss's Zarathustra, measures 5-13 (= same as the reduction in (6), without lower voice) [**AV03a** http://bit.ly/2CR6KMP]

b. Same as a., but removing notes responsible for the downward or upward movement of the soprano in a. [**AV03b** http://bit.ly/2CGU2wk]

c. Same as b., but lowering by one octave the lower Es [**AV03c** http://bit.ly/2mbSHXW]

The potential contribution of the melodic movement can be further highlighted by turning to (ii) [same melodic movement without the harmony] and asking what effect is obtained if we rewrite (8)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. In this case, we have not so much constructed a 'minimal pair' (since there are many differences between (9) and the reduction in (6)) as 'removed' one dimension of the piece, namely harmony. (This is more commonly done when one is interested in the rhythm of a piece without consideration to its tonal properties: one can simply remove the notes.)

(9)   A version of (8)a re-written using only the note C [**AV04** http://bit.ly/2m99bPE]

In sum, both harmonic and non-harmonic properties could conspire to yield a powerful effect in the case at hand, and their potential contributions can be isolated by rewriting the piece in various ways, although this does not tell us what are the respective roles of these two effects. Still, why should one draw such inferences on the basis of loudness and (non-harmonic) pitch height? As a first approximation, we can note that in normal auditory cognition a sound source may 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 (if the amplitude is constant). On the tonal side, normal auditory cognition will not be directly helpful to draw inferences, but it seems that stability properties of tonal pitch space are somehow put in correspondence with stability properties of real world events.

For concreteness, we introduced the issue of semantic inferences using intuitive judgments triggered by a well-known excerpt. But numerous experimental results, referenced below, also establish related facts. Thus Koelsch et al. 2004 show that musical excerpts can prime certain words but not others (e.g. an excerpt might prime 'wideness' rather than 'narrowness', while another does the opposite; and similarly for 'needle' vs. 'river'); furthermore, the brain signatures of this priming effect (N400) are thought to be characteristic of semantic priming (see also Koelsch 2012 for a review). Eitan & Granot (2006) show that "most musical parameters significantly affect several dimensions of motion imagery", while Juslin & Laukka (2003) and Gabrielsson & Lindström (2010) survey numerous emotional effects triggered by various musical parameters.

The challenge will thus be twofold. First, we should argue 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

Sound gives rise to all sorts of inferences about the sources that caused it. 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 (for instance thanks to voice leading principles of classical music theory, and/or by principles of Auditory Scene Analysis applied to music) [13], and as a first approximation we will take the inferences to pertain to the virtual sources of

---

[13] Huron 2016 investigates the cognitive principles (primarily from Auditory Scene Analysis) behind voice leading principles. In his words (from the Introduction), "voice leading is a codified practice that helps

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.[14] We briefly come back to related issues in Section 10 as we discuss the role of emotions in music semantics, but for the most part the present discussion is restricted to very simple effects.

Inferences will be of two general types: some, triggered in particular by timbre and pitch, pertain to what the source *is*; others pertain to what the source *does*, and where it does it (relative to the perceiver): sounds evoke the occurrence of some events, whose speed they reflect; loudness and sometimes pitch modifications convey information about the energy with which the source acts; and sometimes the music just imitates the sounds produced by the source. We do not present the list as closed: if our analysis is on the right track, all sorts of inferential effects found in normal auditory cognition may be recycled in music, and compiling an exhaustive list is not a feasible goal.

4.1   Timbre

While this might be too obvious to state, timbre can give an indication about the identification of the voices, and the sources they correspond to. This is especially true when different timbres 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. A timbre may provide semantic information due to its intrinsic properties: a piano may be less successful than a flute to represent a bird because its sound is less similar to a bird song.[15]

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. This 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 can be found in Saint-Saëns's *Carnival of the Animals* (1886), in the part devoted to kangaroos, illustrated in (10). When the first piano enters, it plays a series of eighth notes separated by eighth silences.[16] This evokes a succession of brief events separated by interruptions. In the context of Saint-Saëns's piece, these sequences evoke 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 far more abstract if we did not have the title and context of the piece, but the main effect would remain, that of a succession of brief, interrupted events.

---

musicians craft simultaneous musical lines so that each line retains its perceptual independence for an enculturated listener".

[14] A similar distinction is needed for non-musical sounds: 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, such as the rain falling or the wind blowing. See footnote 36 for a reference to Leonard Bernstein's discussion of semantic inferences that are arguably licensed by the string accompaniment in Charles Ives's Unanswered Question.

[15] Unsurprisingly, in his *Carnival of the Animals*, Saint-Saëns uses the clarinet to represent a cuckoo [**AV05** http://bit.ly/2mB69Uy] and the flutes to represent an aviary   [**AV06** http://bit.ly/2ELPC7X]. But some semantic effects are more subtle, as in Saint-Saëns's use of flutes in the melody intended to evoke an aquarium [**AV07** http://bit.ly/2D7zkWF]. 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.

[16] As noted by R. Casati (p.c.), the effect is strengthened by the grace note (it might contribute to the understanding of the rebound).

(10) Saint-Saëns's *Carnival of the Animals,* Kangaroos, beginning [**AV08** http://bit.ly/2m98kPd]

Importantly, the inferences one naturally derives from musical events are more abstract than those that normal audition would yield, since inferences may be drawn about virtual sources with no assumption that these produce sound. Our formal account in Section 6 will capture this observation.

4.3   Speed and speed modifications
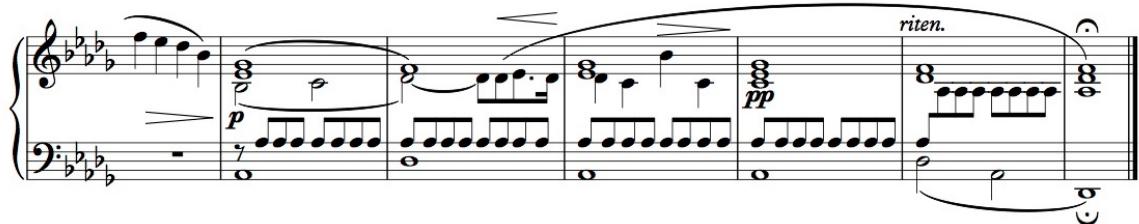
Since sound (as opposed to silence) provides information about events undergone by the source, changes in the speed of musical events will be interpreted as changes in the speed of the denoted events. In the quoted piece on kangaroos (in (10)), each series of jumps starts slow, accelerates, and ends slow. This produces the impression of corresponding changes of 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 [**AV09** http://bit.ly/2DAbnrN]. It features an extremely slow version of a famous dance (the *Cancan)* made popular in an opera by Offenbach (the 'infernal galop'). 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 [**AV10** http://bit.ly/2qM6bhE] 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 [**AV11** http://bit.ly/2B1UkAf] loses much of the solemnity of Mahler's version, and it also sounds significantly happier (a point to which we return in Section 10.2.1).

There are also 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 (see Ilie and Thompson 2006 on the relation between speed and 'energy arousal'). This is probably why it is customary to signal the end of certain pieces with a deceleration or 'final ritard'. An example among many involves 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 does not entirely remove the impression but just weakens it).

(11) Last bars of Chopin's Prelude 15 ('Raindrop')
      a. The last two bars include a ritenuto (normal version). [**AV12a** http://bit.ly/2qHPSmj]

      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. [**AV12b** http://bit.ly/2mcUr2z]

A hypothesis of great interest in the literature is that the precise way in which a final ritard is realized follows laws of human movement within a physical model with a braking force (see Desain and Honing 1996, and Honing 2003, who introduces his idea by way of a mechanical machine that realizes a ritard [**AV13** http://bit.ly/2EKSOAF]).

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, although this would require a separate musical and ethological discussion.

Because of observations of this type, the meaning of music has often been analyzed in connection with *movement* (e.g. Clarke 2001, Eitan and Granot 2006, Godoy and Leman 2010, Larson 2012). But in the general case we will make use of the weaker notion of *change* because music may be interpreted in terms of internal experiences, as we will see in our discussion of musical emotions in Section 10.

4.4   Loudness

A sound that seems to the perceiver to be becoming louder could typically be interpreted in one of two ways: either the source is producing the sound with greater energy, or the source is approaching the perceiver. As Eitan and Granot 2006 write, while "dynamic changes are mostly produced by changes in the energy of the emitted sound", a listener might still "metaphorically relate musical loudness to distance, given a lifelong experience of relating the two features in nonmusical contexts". The first case is of course pervasive in music (for experimental results, see for instance Ilie and Thompson 2006). 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 (12)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.

(12) Mahler's Frère Jacques (First Symphony, 3rd movement)[17]

a. Beginning, normal version [**AV14a** http://bit.ly/2ma7rFW]

Feierlich und gemessen, ohne zu schleppen

 b. Beginning, with an artificially added crescendo: this can yield the impression that a procession is approaching. [**AV14b** http://bit.ly/2m9WnIS]

c. End: depending on the realization, the decrescendo might be indicative of a procession moving away. [**AV14c** http://bit.ly/2mc6oVV]

Without any manipulation, the end of Mahler's Frère Jacques displays a decrescendo which could suggest that the source is gradually losing energy, but which could also be construed as a procession moving away from the perceiver ((12)c).

Interestingly, just by considering 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 first case, one would not expect the perceived speed of events to be significantly affected. In the second case, by contrast, both the loudness and the speed should be affected. The effect can be tested by exaggerating the diminuendo at the end of Chopin's Raindrop Prelude in (11); without the ritenuto, the source is easily perceived as moving away.[18]

(13) Last bars of Chopin's Prelude 15 ('Raindrop')

a. In an exaggerated version of the diminuendo in the normal version, realized with a ritenuto, the source seems to gradually lose energy, becoming slower and softer. [**AV15a**   http://bit.ly/2CJWHVJ]

b. In a version of a. without ritenuto, the source seems to be moving away, as it gradually becomes softer, without change of speed. [**AV15b** http://bit.ly/2qMnRd0]

This type of prediction highlights the importance of a semantic framework that postulates a virtual source behind the music, and simultaneously studies all the inferences it may trigger. In the case at hand, it is because of properties of sound sources in normal auditory cognition that a diminuendo realized with a ritenuto naturally gives rise to an interpretation in terms of gradual loss of energy, whereas a diminuendo without a ritenuto can be interpreted as the source moving way.

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 it licenses about the (virtual) source of the sound. Two kinds of inferences are particularly salient.

(i) The register of a given source – especially if the source is an animal – provides information about its size: larger sources tend to produce sounds with lower frequencies (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).[19] The relevant inference is put to

---

[17] Here and throughout, we follow standard musical convention in notating the contrabass part one octave higher than it sounds.

[18] If we add a crude crescendo instead, and a final accent, the ending sounds more intentional, as if the source gradually gained stamina as it approaches its goal, and signaled its success with a triumphant spike of energy [**AV16** http://bit.ly/2mcPWET]. An intentional, triumphant effect is often produced by fortissimo endings, e.g. at the end of Beethoven's Symphony No. 8 [**AV17** http://bit.ly/2ALmzz2].

[19] 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).

comical effect in Saint-Saëns's *Carnival*, where the melody of a dance is played with a double bass to figure an elephant [**AV18** http://bit.ly/2rbl5OZ]. The specific effect of pitch, keeping everything else constant, can be seen by comparing Saint-Saëns's version (in a MIDI rendition, as in (14)a) to an artificially altered version in which the double bass part was raised by two octaves. The impression that a large animal is evoked immediately disappears. If the double bass part is raised by 3 octaves, a small source is evoked instead (as in (14)c).

(14) Saint-Saëns's *Carnival of the Animals*, The Elephant, beginning
  a. The normal version features a double bass to evoke a large animal. [**AV19a** http://bit.ly/2mea8pQ]
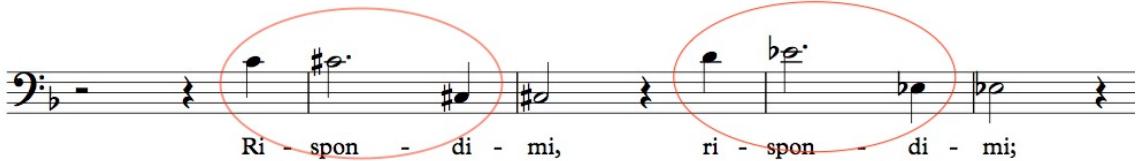
  b. Raising the double bass part by 2 octaves (while leaving the piano accompaniment unchanged) removes the evocation of a large source. [**AV19b** http://bit.ly/2CIOHEp]

  c. Raising the double bass part by 3 octaves might even evoke a small rather than a large source. [**AV19c** http://bit.ly/2CI6Xhk]

(ii) Keeping the source constant, higher pitch is associated with more events per time unit, which suggests that the source might have more energy or be more excited; Ilie and Thompson 2006 provide experimental evidence for an association between higher pitch and greater 'tension arousal' ('tense' vs. 'relaxed'). We already saw an instance of this effect in the version rewritten only with C notes of the beginning of Strauss's *Zarathustra* in (9). 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).

(15) Mozart's Don Giovanni, Commendatore scene (Act II, final scene) 'Rispondimi': repetition is produced with a chromatic ascent, which contributes to the impression that the Commendatore's request is becoming more pressing. [**AV20** http://bit.ly/2ELKqRv]

Ri - spon - di - mi, ri - spon - di - mi;

  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 melodic 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. If the score is re-written so that the piece ends upwards rather than downwards, the effect is arguably a bit less conclusive, as is illustrated in (16).

(16) 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. [**AV21a** http://bit.ly/2CKX0zH]

 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 arguably less conclusive. [**AV21b** http://bit.ly/2Eprsjq]

  Larson 2012 defines a principle of 'melodic gravity' to capture the "tendency of notes above a reference platform to descend" – which comes very close to what an energy-based interpretation of pitches would lead one to expect as a default pattern, i.e. without the intervention of further forces (ones that are analyzed within Larson's theory of 'musical forces'). Similarly, Larson defines a principle of 'musical inertia', which is the "tendency of pitches or durations, or both, to continue in the pattern perceived" (we briefly come back in Section 5 to a further principle of 'melodic magnetism'). Importantly, these are not primitives in the present analysis: when pitch differences trigger inferences about the changing level of energy of a given source, our knowledge of the world will be sufficient to trigger the expectation that, under specific circumstances (and in particular in the absence of external, non-musical forces), the level of energy of that source should go down. Similarly, world knowledge might lead us to expect that, as a default, things might continue to behave as they did (with decreasing energy if 'friction' matters). These effects might be quite real, but on the present view they result from the interaction of music semantics with world knowledge rather than from primitive musical principles: it is because of what we know about the denoted virtual sources that these can be expected to behave in certain ways.

4.6   Imitation

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 [**AV22** http://bit.ly/2FiUum1]. Here timbre, frequency and spatial origin of the sound conspire to produce 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 [**AV23** http://bit.ly/2AJmY4K]. Famously, the Star-spangled Banner is a recurring theme of Puccini's Madama Butterfly [**AV24** http://bit.ly/2B1mrLD], 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 [**AV25** http://bit.ly/2ELiDk1] – with abominable errors – belong to the menagerie described in Saint-Saëns's *Carnival*.

The effects we described in earlier sections 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 heard outside of a musical context: the sound of a cannon is attributed to a virtual source which is a cannon, and a scale with errors can be attributed to a piano student's hapless practice.

4.7   Interaction of properties

Rather than delving more deeply into a topic we must leave for future research, we will give one example that simultaneously involves several factors. 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 [**AV26a** http://bit.ly/2mk9SH3] they could be interpreted in terms of a source moving away, or gradually dying out. But if a strong rallentando is added [**AV26b** http://bit.ly/2miq23k], the 'moving away' interpretation becomes less likely, and the 'dying out' interpretation becomes more salient; this is exactly the effect we discussed in connection with the end of Chopin's Raindrop Prelude in (16).

We can also manipulate the beginning of Mahler's Frère Jacques to modify the interpretation of the initial repetitions. A repetition that is realized far more softly than its antecedent may sound like an echo of it, as in (17)b. A louder realization of the repetition may be interpreted as re-assertion, or possibly as a dialogue between two voices, as in (17)c.

(17) Mahler's Frère Jacques (First Symphony, 3rd movement)
a. Beginning, normal version [**AV27a** http://bit.ly/2AJNAmf]

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. [**AV27b** http://bit.ly/2CVBhcl]

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. [**AV27c** http://bit.ly/2CKoo0O]

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

perceived level of energy may increase. Alternatively, the source may be intentional and may be reiterating an action that was not initially successful, possibly with more energy than the first time around.[20] Yet another possibility is that one source is imitating another. The typology will no doubt have to be enriched.

4.8   Methods

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 with the following method.[21]

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-pitched than a lower-pitched sound.

2. Second, minimal pairs should be constructed to assess the inference in a musical context. This could be done in two ways. One may select actual musical examples, and manipulate them so as to obtain contrasting pairs, as we did with the end of Chopin's Nocturne 9/2 (in (16)). Alternatively, one may create artificial stimuli which also display a minimal contrast with respect to the relevant parameter, but might be simpler than 'real' music, as we did in our discussion of a pure C-version of Strauss's Zarathustra (in (9)).

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 test the target inference by way of 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*? One may also test the inference in indirect ways, for instance by having subjects match musical stimuli with non-musical scenes (e.g. visual ones). Which types of tests will prove most productive is entirely open, and it is likely that different methods will have to be developed depending on the particular goals of the research. Finally, semantic intuitions can be sharpened by initially restricting the set of models the subjects consider. This is 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 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. This may be done by creating non-musical stimuli (e.g. with noise, with human voices, or 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 that case.

5   Semantic effects II: inferences from tonal properties

5.1   Inferences from normal auditory cognition vs. inferences from tonal properties

As mentioned in Section 3.1, Lerdahl 2001 draws an analogy between Heider and Simmel's (1944) animated geometric figures endowed with agency, and semantic effects obtained in music. Importantly, for Lerdahl these inferences arise in part on the basis of the behavior of voices in tonal pitch space. Relatedly, Larson 2012 develops a theory in which the semantic effects of music are analyzed in terms of motion, but within a universe with 'musical forces' that are based in part on harmonic considerations – notably, a principle of 'melodic magnetism', which is "the tendency of unstable notes to move to the closest stable pitch" (p. 2)).

---

[20] In Charles Ives's Unanswered Question, the repetition of the trumpet motive lends itself to a dialogical interpretation: a question is repeated several times in near-identical form, and answers are increasingly frustrating. We revisit this example in Section 9.3.

[21] See Eitan and Granot 2006 for more specific methods designed to test the relation between music and movement.

Since tonal properties do not have a complete equivalent in normal auditory cognition, we must complement our initial list of inferences (from normal auditory cognition) with ones that are specifically drawn on the basis of tonal properties. The challenge (to be addressed in Section 6) will be to develop a method to aggregate these heterogeneous inferences.[22]

Tonal pitch space comes in different varieties in different musical traditions, and even within a given musical tradition, as shown by the distinction between major and minor keys. Correspondingly, inferences drawn on the basis of the behavior of the voices in tonal pitch space will depend on the musical idiom under study, and they should thus not be expected to be invariant across cultural traditions.

As mentioned at the outset, the meaning of a musical piece is sometimes equated with a journey through tonal pitch space, as is informally suggested by Lerdahl 2001 and formally implemented in Granroth-Wilding and Steedman 2014. Within this 'tonal journey' direction, one is sometimes tempted to reduce music semantics to a model of musical tension, as developed for instance by Lerdahl 2001 and Lerdahl and Krumhansl 2007. Musical tension is indeed crucial to music semantics, but it doesn't follow from this that musical meaning *reduces* to musical tension. Rather, the sources of musical events are understood to be located in a space which is isomorphic to (or at least shares some formal properties with) tonal pitch space, and it is for this reason that the relative stability of these positions, and the attraction relations among them, are essential to understand the events undergone by the sources. It is thus crucial to aggregate inferences from tonal properties with inferences from normal auditory cognition, as we proposed to do in this piece.

The rest of this section motivates the existence of specifically tonal inferences. Section 6 will then sketch a 'toy model' in which inferences from normal auditory cognition and tonal inferences can be aggregated.

5.2   An example: a dissonance

A very simple example will help illustrate the inferential power of tonal inferences. In Saint-Saëns's very slow version of the *Cancan* dance, which he uses to represent tortoises, there are moments of severe dissonance, and they produce a powerful effect. The very slow dance evokes the tortoises' slow walk. But when we hear a dissonance in measure 12, circled in (18), 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". While at first it may seem that the musicians are out of tune, in fact they are just playing a dissonant chord, with both A and G# in the same chord, as shown in (18). When the G# is replaced with A throughout this half-measure (as in (18)b), the dissonance disappears, as does the impression that the tortoises are tripping.

---

[22] We should note that while tonal inferences can only be understood by reference to the formal properties of tonal pitch space, they might well be *grounded* in some properties of normal auditory cognition, for instance in animal signals, human voices, or more general inferences relating consonance/dissonance to properties of the source; we briefly discuss some possibilities at the end of this section.

(18)   Saint-Saëns, *Carnival of the Animals*, Tortoises, measures 10-13 [**AV28** http://bit.ly/2DAeq3d]

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 G# added (as can be heard by focusing only on the violin and piano parts). [**AV28a** http://bit.ly/2ECNWNJ]

b. The dissonance can be removed by turning the G#'s into A's – and the impression that tortoises disappears (as can be heard by focusing only on the violin and piano parts). [**AV28b** http://bit.ly/2CWFVCT]

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, the intended virtual source. In the general case, things are far less specific. In fact, if we disregarded Saint-Saëns's title, the inferences we draw would not specifically be about tortoises, but they would still probably involve a source which is slow (due to the comparison with the speed of the standard *Cancan*), and also goes through positions of instability at moments that correspond to the dissonances (this would be *compatible* with the tortoise-related interpretation, but far less specific).

5.3   *Cadences*

In traditional music theory, 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' ending on a dominant chord, which can signal temporary pauses and call for a continuation. These devices play a central role in analyses of musical syntax, as in Lerdahl and Jackendoff 1983, and Rohrmeier 2011 (for whom cadences play a crucial role in the generation of syntactic trees by way of rules of 'functional expansion').

The question that is not fully addressed in these syntactic frameworks 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 might need to be stated within a semantic framework. In brief, a full cadence is final because it ends in a position of tonal 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, assuming that the hierarchy of stability of chords is IV < V < I; this 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 could in principle capture these facts as follows: music is special (compared to non-musical sounds) 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 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.

Of course this only scratches the surface of an analysis of cadences. Still, the general form of the account seems appropriate to account for more fine-grained phenomena. To mention just two:
• A cadence is more conclusive if the final tonic chord in root than 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 two 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'.

It is worth giving an example of the effect of the slightly 'incomplete' feeling produced by a deceptive cadence. (19)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 (19)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 less conclusive.

(19) Ah vous dirai-je Maman, simplified from Mozart's theme (b. was written by A. Bonetto)
    a. Perfect cadence: II V I [AV29a http://bit.ly/2DohwYa]

    b. Deceptive cadence: II V VI [AV29b http://bit.ly/2D7fMEI]

For concreteness, we have focused on a particular excerpt rewritten in various ways. But rich experimental results exist as well. Thus Rosner and Narmour 1992 systematically assessed the relative closure of chord progressions in naive subjects. They found clear differences across chord types, with V-I sequences assessed as more closed than all other progressions, in particular III-I, VI-I, or the plagal cadence IV-I. Progressions were generally assessed as more closed when the root was in bass position. Thus the general claims of traditional music theory seem to be empirically legitimate.

While the topic of cadences is a staple of music analysis, which the foregoing remarks just recapitulate, we believe that they should be studied within a broader framework in which considerations of harmonic stability are investigated 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 modifications trigger different inferences, and that they can be combined to suggest 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 reached, interact with the inferences triggered by loudness and speed, pitch, rhythm, etc. This is certainly not a new idea – for instance, Heinrich Schenker's influential theory took not just harmonic progression but also a melodic 'fundamental line' to be part a complete tonal piece (e.g. Forte 1959, Pankhurst 2008).

5.4 *Modulations*

As is also well-known, tonal pitch space is organized into regions, which correspond to keys – with relations of distance among those. Modulation is often discussed through the metaphor of a movement towards a new location (Saslaw 1996), which may be more or less distant depending on the nature of the modulation (Thompson and Cuddy 1992 provide evidence that listeners with moderate musical training are indeed sensitive to the distance between keys in modulations). While experimental

evidence would be needed to establish this point, we submit that moving to another key triggers the inference 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 source: 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 in G Major but modulates to B minor in measures 7-10, as seen in (20)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 (20)b,c.

(20) 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. [AV30a http://bit.ly/2D6TcNq]

    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. [AV30b http://bit.ly/2DqCC80]

    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 Major. [AV30c http://bit.ly/2ED4yVY]

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.

5.5   Methods and further questions

Having sketched some very simple semantic effects that are triggered by tonal properties of music, we should add a word about the methods that could be employed to investigate them. 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 tonality is not found in non-musical sounds, part (ii) is not applicable in the present case. Thus the analysis must *per force* be more theory-internal. We propose that it should include the following steps.

1. First, a hypothesis should be stated – for instance that a dissonance can trigger the inference that the source is in unstable position (as in our discussion of Saint-Saëns's Tortoises in (18)).

2. Second, minimal pairs should be constructed to establish the point. As in the case of inferences from normal auditory cognition, 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 (this is precisely what Saint-Saëns's titles *The Swan* or *Tortoises* do in the cases we just discussed).

3. Third, instead of correlating these effects with ones that are found in non-musical stimuli, one can seek to explain them by properties of tonal pitch space as analyzed (on non-semantic grounds) by the best experimental and formal studies available.

Still, although some of the key properties of tonal pitch space are not commonly found in normal auditory cognition, one could ask whether normal auditory cognition motivates some of the general inferences we draw on the basis of tonal pitch space. We argued 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. But what is the basis for this general inference? It would be interesting to investigate inferences produced by highly dissonant sounds 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 that one could somehow do without the properties of tonal pitch space in stating a music semantics). This enterprise would require an understanding of the acoustic basis of consonance and dissonance, which has been studied in detail (e.g. McDermott et al. 2010), and also of its correlates in the natural world.

It must be mentioned, however, that the experimental literature usually focuses exclusively on the connection between tonal properties and *emotions* (a topic we revisit in Section 10). For instance, Bowling et al. 2010 compare American speech and music, and write that "the spectral characteristics of excited speech more closely reflect the spectral characteristics of intervals in major music, whereas the spectral characteristics of subdued speech more closely reflect the spectral characteristics of intervals that distinguish minor music" (see also Bowling et al. 2012). For his part, Cook 2007 argues that the emotional effect of minor vs. major chords is related to Ohala's 'frequency code' (e.g. Ohala 1994), according to which animal dominance is expressed with low and/or falling pitch (Cook's proposed connection is that "tension triads resolve to minor chords with a semitone increase and to major chords with a semitone decrease", and "pitch decreases connote positive affect and pitch increases connote negative affect"). Going in a somewhat different direction, Blumstein et al. 2012 show that adding distortion noise (nonlinearities) in a musical piece induced in listeners an effect of "increased arousal (i.e. perceived emotional stimulation) and negative valence (i.e. perceived degree of negativity or sadness)". It is thus fair to say that the direct connection we propose to establish between tonal stability and the stability of *external events* denoted by the music has yet to be tested empirically.

6 Musical truth

We showed in Section 4 that diverse semantic inferences are drawn in music from properties of normal auditory cognition. We saw in Section 5 that further inferences are drawn on the basis of properties of tonal pitch space. We will now sketch a formal framework in which these two inferential types can be integrated.

This enterprise matters for three reasons. First, the inferences we displayed are abstract, and one must state precisely how they are drawn. For instance, in our discussion of Saint-Saëns's Kangaroos, we argued that a source-based semantics can explain why a series of eighth notes separated by eighth silences can evoke a succession of brief events separated by interruptions. But certainly our source-based semantics should not lead to the absurd inference that *kangaroos* are producing these notes – or sounds, for that matter. Rather, something more abstract is inferred from the music, namely that there was a quick succession of discrete events; all sorts of events, whether sound-producing or not, will satisfy this abstract inference. Second, the inferences we discussed interact with each other in non-trivial ways. As we saw in Section 4.7, a repetition with attenuation may be interpreted as a source dying out or moving away, but the former interpretation seems to become more likely when a rallentando is added. The key is that objects that move away without losing energy are unlikely to slow down, contrary to objects that are losing energy. We must thus find a systematic way to integrate inferences with one another, and also with world knowledge. Third, a systematic framework for musical inferences will turn out to yield a natural notion of 'musical truth', which is of interest in its own right.

6.1 Inferences and interpretations

In view of the existence of inferences from normal auditory cognition as well as from tonal properties, the main challenge is to define a framework that can aggregate them despite their heterogeneity.
In principle, this could be done in two ways:
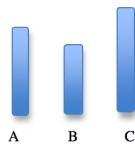
1. **Inferential direction:** we could find a way to simply conjoin all the relevant inferences – and say that *the meaning of a musical piece is the set of inferences it licenses on its sources*.

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 (or 'model').
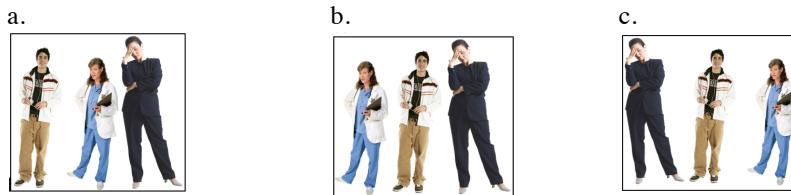
An advantage of the second method is to ensure that the inferences licensed are not contradictory: by providing a situation that makes all of them true, we can be sure that we are not dealing with a system that is trivial because it licenses contradictions.[23] Still, it is often more intuitive to speak 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. In addition, as we will see in Section 6.3, the definition of a notion of musical truth makes it possible to obtain a derived notion of *semantic content* for a musical piece.

Under what conditions is a musical piece true of a situation? We will take musical events to depict events undergone by virtual sources. And as a first approximation, we will take a series of musical events to be true of a series of world events if certain relations among notes or chords correspond to designated relations among events; for instance, a louder note should correspond to a world event which has greater energy or is closer to the perceiver; a more consonant chord should correspond to a more stable world event, etc. The basic mechanism can be illustrated in a different domain by considering simplified pictorial representations, seen as visual depictions of certain objects. An example is given in (21), where three columns of various heights (A, B, C), arranged from left to right, are used to depict individuals as in the scenes in (21), involving a boy, a nurse and business woman.

(21) A pictorial representation

(22) Three possible denotations for (21)

a.   b.   c.

We focus on two relations among the columns that appear in (21): 'is to the left of' (from our perspective), and 'is taller than'. At a very coarse-grained level, we can say that an assignment of values (namely real world individuals) to the columns makes the picture *true* in a certain scene if these two relations are preserved.

Consider the assignment A → boy, B → nurse and C → businesswoman in the scene (21)a. A is to the left of B, which is to the left of C; the same relations hold of the denotations in the scene, since the boy is to the left of the nurse, who is to the left of the business woman. Thus the relation 'is to the left of' is preserved. Similarly for the relation 'is taller than': C is taller than A, who is taller than B. The same relation holds of the denotations, since the businesswoman is taller than the boy, who is

---

[23] We set aside the case of auditory illusions with a contradictory content.

taller than the nurse. Thus we can say that on this assignment of values to the columns, the pictorial representation in (21) is true of (22)a. By contrast, it is immediate that the assignment A → nurse, B → boy and C → businesswoman would fail to preserve the relation 'is to the left of', since (from our perspective) A is to the left of B in (21), but the nurse is not to the left of the boy in (22)a.

By similar reasoning, on the assignment A → nurse, B → boy and C → business woman, the relation 'is to the left of' in (21) is preserved in scene (22)b. But the relation 'is taller than' is not preserved: while A is taller B, the denotation of A, the nurse, is not taller than the denotation of B, the boy, hence on this assignment (21) is not true of scene (22)b. In fact, no assignment of denotations could preserve both 'is to the left of' and is 'taller than' in this case, and similar remarks hold for (22)c.

We will apply the same type of definition of truth to musical pieces, but with relations that are more abstract than those involved in this simple pictorial example. Since musical pieces are dynamic, something like the relation 'is to the left of' will be played by the relation 'temporally precedes': we will require that the denoted events appear in the same order as the notes that represent them. We will also add further preservation principles that will play the same kind of role as height preservation; for instance, we will require that a more stable chord should refer to a more stable event.

In our pictorial example, one may well investigate more fine-grained conditions of preservation, for instance involving the *proportions* among columns rather than just the relation 'is taller than'. Similar refinements should be investigated in the musical case, but here we will be content to sketch the barest of semantics in order to provide a 'proof of concept', leaving such refinements for future research.

6.2 *An example of musical truth*

Because what precedes is rather abstract, we should start with a highly simplified example. Let us consider again 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. First, 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. Second, the progression is realized with a 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 (23) (with loudness expressed in decibels, dB). 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).

(23) 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 a diminuendo progression from IV to V to I. For present purposes, a musical piece is just an ordered series of such pairs. Those 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 (there is of course no requirement that the denotations should be *actual* events, i.e. events that did or will in fact happen[24]: just as pictorial representations, music can be fictional). For maximum simplicity, our musical pieces will be reduced to a single voice. Each such piece/voice will include three musical events, as illustrated in (23), which will depict a series of 3 possible events in the world. But as we saw earlier, events are not enough: inferences are derived by considering virtual sources of the voices, and these sources are often identified with possible *objects in the world*. Accordingly, we associate:

---

[24] We sometimes contrast 'real world events' with 'musical events', but in all cases our world events are possibilia.

(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 (possible) world events $e_1, \ldots, e_n$, with the requirement that each of these events should have O as a participant.

This is made precise in (24).

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

(See Wolff (2015) for a rather different event-based analysis of musical meaning, one without a notion of 'musical truth'.)

The next step is to determine under what conditions a series of musical events can be taken to be true of world events. In our analysis, this will be the case when these world events satisfy certain inferences triggered by the musical voice – inferences from normal auditory cognition, and tonal inferences. Here we will only give a toy example of an analysis of this kind: our goal is merely to illustrate the conceptual points we are making, and we will leave it for future research to develop analyses that are more realistic and thus take into account more parameters as well as more preservation principles.

We start from pieces such as those in (23) (each reduced to a single voice), combined with the specification of possible denotations in (24). We will say that the musical piece M = $<M_1, \ldots, M_n>$ (made of n musical events) is true of the pair of an object and events it participates in, $<O, <e_1, \ldots, e_n>>$, 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 (25).

---

(25)   Defining 'true 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 $<O, <e_1, \ldots, e_n>>$** 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 does not require justification, the Loudness and Harmonic stability conditions do. Let us consider them in turn.

The preservation condition on Loudness is disjunctive. The intuition is that in auditory cognition in general, louder sounds are associated either with objects that have more energy, or with objects that are closer to the perceiver, 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.

Two essential remarks should be added. First, *none of the conditions in (25) require that the denotations produce sound*. This is the sense in which our source-based semantics is abstract: the properties we attribute to the objects are ones that would be inferred about sound sources, but these properties themselves need not involve sound, and thus they may be true of objects that are not sound-producing. Second, a musical piece will in general be true of numerous objects and their associated events. The same situation arises in most semantic systems, such as human language: to understand the meaning of the sentence *It is raining* is to know in which kinds of situations it is true, but the sentence need not refer to a single situation.[25] Still, it is particularly striking that in music the denoted

---

[25] See for instance Larson (1995) and Schlenker (2010) for handbook summaries of the analysis of linguistic meaning as truth conditions.

situations may be extremely heterogeneous, as we will see shortly. This is because the informational content of music is underspecified and abstract, which has led some to think that music has no semantics at all. But an underspecified and abstract semantics is very different from no semantics at all.

We can now illustrate how these preservation conditions can deliver a notion of truth. We consider three objects: the sun, a boat, a car. And we will consider 'bare bones' versions of several sequences of possible events. For the sun, a sunrise and a sunset. For the boat, a movement towards the perceiver, and a movement away from the perceiver. 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 (23)a.

(26) 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, <minimal-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 (26) is of the form <object, <event_1, event_2, event_3>>, each is a possible denotation for M according to (24). 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 (I-V-I); and loudness increases, which can be interpreted as a rise in (real or perceived) level of energy, as in Sun-rise, or as an object approaching, 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 world events on independent grounds.

Let us first note that all the sequences of events given in (26) are intended to obey the time ordering condition stated in (25)a: 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 (26), all we need to check is that it satisfies the Loudness and the Harmonic Stability conditions.

• Consider first Sun-rise in (26)a. Since M has a crescendo, M_1 is less loud than M_2, which is less loud than M_3. The Loudness condition in (25)b mandates that minimal-luminosity should have less energy or be further from the perceiver than rising-luminosity; and similarly 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, but 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 sub-events 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 (26)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.

• The argument is almost identical in (26)c,d as in (26)a,b (in particular with respect to the Harmonic Stability condition), but with one difference: since it does not make much sense to say that a boat approaching is gaining energy (if anything, it might slow down as it approaches the coast), the

Loudness condition is satisfied in (26)c by an increasing proximity of the source to the perceiver (fulfilling (25)b(ii)) rather than by an increasing level of energy of the source (pertaining to (25)b(i)). The Loudness condition is violated in (26)d: its last two sub-events are departure followed by maximal-distance, and the second does not 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 (26)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 violated: 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 tonic (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. Needless to say, neither the sun nor the boat need to produce sound in order to be denoted, which we take to be an appropriate result, and a benefit of the formal approach sketched here (without it, one might think that a source-based semantics can only posit sound-producing denotations, which would be undesirable). In the general case, a piece will likely be made true by extremely diverse situations, because our preservation conditions make reference to abstract properties (e.g. level of energy, stability) that could be instantiated in countless ways. This is as it should be: musical inferences are highly underspecified, and this property should be preserved by an adequate semantics. From the present perspective, to understand the meaning of a sequence of notes is to understand which possible denotations make it true (which does not imply fixating on any specific one of these denotations). This understanding may be sharpened by extrinsic considerations (in addition to world knowledge), such as titles in program music, or extra-musical considerations in dance and opera: these may be taken to reduce the set of possible denotations that make the music true. But as is the case for language, there will in general be a multiplicity of situations that make a piece true.

6.3   Truth and semantic content

It is standard to use the truth conditions of an expression to define its semantic content. For instance, once one has defined the truth conditions of *It is raining*, one may take its content to be the set of situations that make the sentence true, and thus the set of situations in which it is raining. The same move can be made in the present framework. In a nutshell, the semantic content of a musical piece can be identified with the set of objects and associated events it is true of. This is defined for the special case of a single voice in (27).

(27) Let $M = <M_1, …, M_n>$ be a voice. The semantic content of M is the set of pairs $<O, <e_1, …, e_n>>$ (where O is an object and $e_1, …, e_n$ is a series of n events) such that M is true of $<O, <e_1, …, e_n>>$ (according to the definition in (25)).

Some clarificatory remarks should be added, pertaining both to the definition of truth in (25) and to the definition of content in (27).
1. As already emphasized, M may be true of very diverse objects and events: there is no requirement that the content should be relativized to a single object.
2. The theory does not place limitations on the types of objects that M could be true of: they could be taken to be real objects, possible objects, Platonic entities, etc.
3. Our talk of distance from the perceiver in (25)b(ii) implies that our analysis is implicitly relativized to a perspective. For simplicity, we can take the perceiver to be given once and for all, but in a more general treatment one might relativize both the definition of truth and the derived notion of content to such a perspectival point (see Lewis 1979 for a similar move for thoughts, and Schlenker 2011 for a survey of related issues in linguistic semantics).
4. Besides extra-musical information such as titles, plausibility considerations will help reduce the set of situations that are denoted by a piece. In particular, the inferential means that are lifted from normal auditory cognition are likely to inherit some its specific properties. For instance, we noted earlier that constant speed combined with decreasing loudness at the end of a piece is likely to be

interpreted as the source moving away. This is presumably because this combination of properties in normal auditory cognition is often due to a similar movement of the source. We leave it open how further reasoning-based considerations could interact with the present framework.

6.4 Model-theoretic truth vs. inferential truth

The toy example of Section 6.2 was developed in order to illustrate 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 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 extra-musical events.

This last step could be taken in two ways. One possibility is to proceed in an inferential fashion: one takes the set of all entailments that can be stated in terms of loudness relations or harmonic stability relations on the musical side, and one reinterprets them in terms of energy/remoteness and event stability. Thus one can observe in the case of M in (23)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 this way, we reinterpret with 'real world' vocabulary some musical relations that involve 'musical' vocabulary pertaining to loudness or harmonic stability. Proceeding in this inferential manner, we can 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 event stability). However, this procedure comes at a cost: when one requires that a set of propositions should be true together, one is not assured that these are not collectively contradictory. To show that they are not, one must find a model that satisfies them all. Precisely this result is delivered by the model-theoretic analysis we sketched 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 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 7.2), 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.

7 Comparisons: logical semantics and iconic semantics

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.

7.1 Differences between music semantics and logical semantics

Our music semantics is entirely different from a standard logical semantics. To see this, it might help to define a very simple logical system in which all sentences are concatenations of propositional letters, and thus of the form p_i, $p_i p_k$, $p_i p_k p_r$, etc. The syntax is similar to what would be obtained with concatenations of notes. We could try to make the semantics as close as possible to that of our music semantics by taking these propositional letters to be true of events, and by adding that concatenation is interpreted as conjunction. In this way, $p_i p_r$ is true of those events that make true both p_i and p_k, and by the same token the sequence $p_i p_k p_r$ is true of those events that make true p_i and p_k and p_r (a more precise definition of this semantics is given in Appendix II).

In this way, one can think of $p_1 p_2 p_3$ as a series of musical events, which may be true of some events. But the similarities with our music semantics end there. First, this logical system has no counterparts of our preservation principles (Time, Loudness, Harmonic stability); 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, an event satisfies $p_1 p_2 p_3$ just in case its satisfies each of the propositional letters $p_1 p_2 p_3$, whereas in our music semantics, a separate subevent is denoted by each note/chord. Third, and relatedly, when we combine two atomic letters of our conjunctive logic, the order in which they are combined is irrelevant to the meaning of the result. This is very different from the case of music semantics, where we took the sequence of musical events to be dynamic representations of world events, with the result that the order in which the musical events appear crucially affects the resulting meaning.

7.2   *Similarities between music semantics and iconic semantics*

A better point of comparison for music semantics can be found in dynamic visual representations such as films or iconic gestures and iconic signs. We discussed at the outset the relevance for music semantics of Heider and Simmel's abstract animations, in which geometric 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* in the sentence in (28). The verb can be realized in a variety of ways, six of which are represented in (29). The second row represents different realizations of the slow version of the sign, with the beginning of the sign in the top picture and the end of the sign in the bottom picture, and the meaning obtained; it is clear that *the broader the end points of the sign, the larger the final size of the group*. The third row represents different realizations of the fast version of the sign (without pictures, as these would be rather similar to those of the slow version), with their meanings as well. The relevant observation is that *the more rapid the movement, the quicker the growth process*.[26]

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

(29)   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

Formally, two properties of the sign are preserved by semantic interpretation, as stated in (30).

(30)   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.

---

[26] The paradigm was not fully minimal, in the sense that further aspects of the sign tended to be modified as well. For more controlled paradigms, see Schlenker, to appear.

b. Speed condition
If $GROW_i$ 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, these preservation conditions bear a formal resemblance to those we posited in the Loudness and the Harmonic stability conditions of 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 has a lexical meaning (stored in memory) which specifies that it is true of events of growth (Davidson 1967). Because this verb *also* has an iconic life, its meaning is enriched by the preservation requirements in (30). By contrast, in our music semantics there is no lexical meaning whatsoever, and the action lies entirely in the iconic principles.

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.[27]

Finally, a terminological issue should be mentioned. As a first approximation, we took musical events to have meaning *qua* Peircian indices (because they involve a causal connection between a signal and its source), rather than *qua* icons (which would involve a resemblance between a sign and its denotation). But the technical theory developed in Section 6 is based on certain preservation conditions that can qualify as 'iconic'. So is our music semantics based on iconicity? It depends on how iconicity is understood. If it involves a kind of intuitive *resemblance* between the signal and its denotation, our semantics need not be iconic. For instance, we took the beginning of Strauss's Zarathustra to be true, among others, of a sunrise. But a sunrise is a silent event that doesn't much resemble a musical piece. On the other hand, if the notion of iconicity is made more abstract, the preservation principles we introduced in our formal analysis (in Section 6) do qualify as iconic: a sunrise could be denoted by the Strauss passage because the mapping between the relevant series of notes and the relevant series of subevents satisfies pre-determined preservation principles. There is thus a terminological point that might require further conceptual elaboration.

8   The syntax/semantics interface

8.1   *Goals*

In any system that has a syntax and semantics (including English), one must ask about their interaction or 'interface'. This includes two types of questions. First, should a given contrast receive a syntactic or a semantic explanation? The intuitive deviance of *John admires herself* is arguably semantic: there is a gender mismatch between the proper name and the reflexive because we assume that *John* denotes a male individual. By contrast, *Admires John himself* is weird for a syntactic reason: the words appear in the wrong order. Second, for sequences that are acceptable, how is the semantics read off the syntax: does it just involve the surface word order or, as is commonly assumed for language, derived from a more abstract tree structure? We shall now address both questions in turn: we will suggest that some of the structural effects that are usually attributed to musical syntax (in Lerdahl's and Jackendoff's framework) might have a semantic origin; and we will briefly explain how something like the present semantic analysis could be articulated with Lerdahl and Jackendoff's syntax.

Our primary goal is to argue that the 'grouping structures' postulated by Lerdahl and Jackendoff 1983 derive from an attempt to organize the musical surface in a way that preserves the structure of the denoted events (we take this interpretation to be in the spirit of Lerdahl and Jackendoff, who emphasize that grouping principles come from perception rather than from rules of a generative syntax). In particular, we will propose that a musical group A is taken to belong to a musical group B if (on any true interpretation) the world event denoted by A can naturally be taken to

---

[27] 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* [AV31 http://bit.ly/2EKq1vT], 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.

be a sub-event of that denoted by B. In other words, grouping structure will be taken to reflect the 'part-of' relations among the denoted events, what is called 'mereology' (or sometimes 'partology') in semantics. We will speculate that this semantic approach might even extend to Lerdahl and Jackendoff's 'time span structures'.

Three clarifications will be useful at the outset. First, we emphasized in Section 6 that our analysis is appropriately abstract: although the properties assigned to possible denotations are ones that would be inferred about sound sources, these properties themselves need not involve sound, and thus they may be true of objects that are not sound-producing. Still, the principles by which we structure the music may stem from general principles by which auditory stimuli are sequenced so as to correspond to the structure of the events that caused them. The situation is in this respect reminiscent of visual diagrams used to represent non-visual stimuli. For instance, although the graph in (7)c represents sound (specifically, loudness) rather than visually perceptible objects, we naturally sequence it using general principles of visual perception *as if* we were trying to uncover the structure of objects that caused this visual stimulus.

Second, the analysis we are about to develop takes the tree-like structure of musical syntax *not* to be of the same nature as that found in linguistic syntax. Conceptually, tree structures in linguistic syntax are often taken to reflect the way in which words are put together (this is sometimes called their 'derivational history': in several theories, tree structures just reflect the derivational history of sentences[28]). By contrast, we take the musical syntax under consideration here to stem from the fact that *auditory stimuli are usually structured so as to reflect the structure of the denoted events*. Technically, following Lerdahl and Jackendoff 1983, we will take the tree structures obtained in this musical syntax to be less constrained than standard 'derivation trees' in linguistic syntax.

Third, we agree with much formal work (including Lerdahl and Jackendoff 1983) in taking musical structure to be a mental construct. But instead of taking it to be produced by a separate syntactic module, we will seek to derive some of its properties from the perceiver's attempt to recover the structure of the denoted events.

8.2   *Levels of musical structure*

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.

Some of Lerdahl and Jackendoff's structures have been analyzed in terms of a generative syntax, as was done by Pesetsky and Katz 2009 for prolongational reductions. By contrast, in most of this discussion we will be solely concerned with grouping structure and time-span reductions. Lerdahl and Jackendoff's own theory departs in two respects from a 'generative syntax' analysis.
(i) First, they take their structures to be based on parsing rather than on generation, and to rely heavily on preference principles rather than on 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 2009 argue. But (ii) is 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*.

---

[28] See for instance Sportiche et al. 2013 for a textbook introduction.

8.3 Grouping structure and event mereology

Grouping structures, as we will now argue, 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 is 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 tree-based analysis would yield, in particular because in some cases it allows for overlap among groups. Third, we will refer to literature on event perception that suggests that events are indeed perceived as structured.

8.3.1 Event mereology and tree structures

Events are standardly analyzed as having a part-of structure, with large events being made of smaller events (e.g. Varzi 2015). 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 (31); it has in particular the consequence that if an event e has parts, then *their* parts are also parts of e (Transitivity).

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

The notion of 'proper part' follows from that of 'part': x is a proper part of y if and only if (henceforth: 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 (32).

(32) 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.

(33) 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 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 (34).

(34) If x and y are atomic events, either x < y or y < x, where < is a temporal ordering.
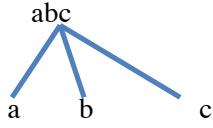
We henceforth use the list of its atoms to name an event, omitting 'trivial' decompositions, namely those that involve events with just two atomic parts (since these can be decomposed in just one way). For an event with atomic sub-events *a*, *b*, *c*, this leads to the possible decompositions in (35).

(35) 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**
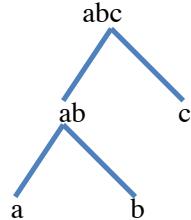
Now it can immediately be seen that (35)a,b,c correspond to 'standard' 'syntactic' trees that could be obtained from a context-free grammar, as illustrated in (36)a,b,c. But (35)d,e require 'trees' with an unusual shape, as illustrated (36)d,e.

(36) a. abc / a b c

b. abc / ab c / a b

c. abc / a bc / b c

d. abc / ac / a b c

e. abc / ab bc / a b c
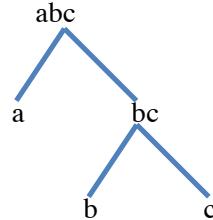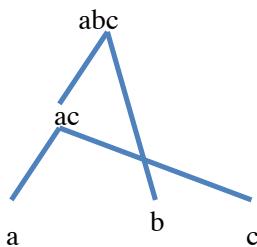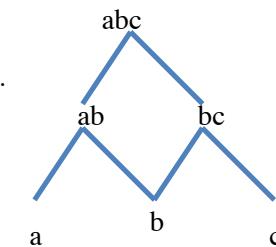
The situation in (36)d violates the assumption that 'constituents are not discontinuous' (a standard but not universal assumption in linguistics, see e.g. McCawley (1982) for exceptions). In standard syntax, it is normally prohibited by the assumption that in a context-free rule of the form M → $D_1 \ldots D_n$, the output elements $D_1 \ldots D_n$ are temporally ordered, with $D_1 < \ldots < D_n$, and 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 (36)d, as we can neither have $ac < b$ nor $b < ac$.

The situation in (36)e 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 (36)e. It is an uneconomical event decomposition, because we could remove a branch above *b* (thus attributing *b* exclusively 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 (37), which prohibits overlap among events unless one is contained within the other.

(37) 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 violated by (36)e: *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 (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 [this might be relevant for modulations]);
(ii) occlusion: cases in which there is a reason to think that two distinct events share the same auditory trace.
We argue in Appendix III-A that precisely these two cases arise in Lerdahl and Jackendoff's analysis of musical syntax. In other words, the mereology-based reconstruction of musical syntax has the advantage of predicting some cases in which musical structures are less constrained than tree structures.

Consider now (36)d. It leads one to posit that an event has a discontinuous auditory trace. Two assumptions are needed to prohibit this case.

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 a single underlying object or event despite the disconnected nature of the percept.

Thus in order to prohibit structures such as (36)d, we must posit that cases of auditory occlusion do not 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 that the background noise is so loud as to fully occlude the conversation, or conversely.

In this case as well, we expect that there should be exceptions, of two types; whether these arise in music has yet to be investigated.

(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.

8.3.2   Event structure

For our analysis to be plausible, we would need to establish that *independently from music* (or language, for that matter), events are naturally perceived with a part-of structure. Jackendoff 2009 argues that there are tree-like structures outside of language, and he gives the example of actions, which may be structured in various ways without thereby having a linguistic representation. In the experimental literature, Zacks et al. 2001 provide evidence that subjects sequence events (presented by way of videos) in a hierarchical fashion. And work by Neil Cohn (e.g. Cohn et al. 2014) suggests that visual narratives (comics) have a hierarchical structure as well. In the future, it would be particularly interesting for music semantics to investigate cases in which two events may overlap, something which is crucial to our understanding of Lerdahl and Jackendoff's cases of grouping overlap.[29]

8.4   Time-span reductions and headed events

We briefly turn to the interaction between musical meaning and time-span structures, which play an important role in Lerdahl and Jackendoff's syntactic analysis.

Lerdahl and Jackendoff argue that their grouping structures are insufficient in that they fail to distinguish different levels of importance within musical groups. They propose that their tree structures are *headed*: at each level, each group contains a musical event that is more important than the others and thus counts as its 'head'. In a nutshell, heads are events that are rhythmically more prominent and/or harmonically more stable. Metrical structure (= the alternation and weak and strong beats) helps select the most important notes at micro-levels, as is illustrated in (38). At larger levels, heads of musical groups are selected by a combination of metrical and harmonic considerations. Thus one can derive from a metrical and grouping structure as in (38) a time-span structure as in (39),

---

[29] An essential issue for future research will be to determine *to what extent the details of musical meaning affect musical structure*. One could adhere to the remarks made above by taking them as a simple reconstruction of Lerdahl and Jackendoff's Gestalt-based views. On this deflationary view, Gestalt principles of grouping arise from an attempt to recover the structure of the *actual* events that caused an auditory percept, and no reference to fictional sources is needed. But it could also be that the details of our semantics affect grouping structure. As an example, one could imagine that in a sequence > < (diminuendo followed by a crescendo), one will be more tempted to put a group boundary after > if it is realized with a strong rallentando, as this suggests that the source is dying out, and that the following crescendo (<) corresponds to a very different event and possibly to a different source. Such issues have yet to be investigated.

where certain chords (notated with Roman numerals) are represented as the heads of the various groups.[30]

(38) Metrical structure [square brackets] and grouping structure [round brackets] for the beginning of Mozart's K. 331 piano sonata (Lerdahl and Jackendoff 1983) [**AV32** http://bit.ly/2DamRom]

(39) Time-span reduction obtained from (38) by selecting in each the musical event which is metrically strongest/harmonically most stable (Lerdahl and Jackendoff 1983)

It remains to ask whether the headed nature of time-spans should be taken as primitive, or might follow instead from a more general strategy of event perception. Jackendoff 2009 argues that there are headed structures outside of music and language, in particular in the domain of complex action. 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 often perceived not just as structured but also as headed, and (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 one suggestive example. Consider a simplified dynamic representation of a person walking, as in (40). We submit that if one were to sequence the walk into events and sub-events, one would find that moments at which the foot touches the ground delimit events, but in addition that these are the most important sub-events in each cycle – the 'heads' of the relevant events, in terms of the present discussion. These are clearly points at which impulses of energy are given, somewhat like points of metrical strength in music, and probably also points of greatest physically stability.

---

[30] As an example, "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 time-spans at that level)." (Lerdahl and Jackendoff 1983 p. 120)

(40) Person walking[31]

It should be added that Lerdahl and Jackendoff take another notion of structure, prolongational reductions, to play a central role in music perception; some questions they raise in connection with music semantics are stated in Appendix III-B.

8.5   Structural interpretive rules?

We speculated in the preceding section 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 (41), where $[[\bullet]]$ is the interpretation function, which assigns to a musical event $\bullet$ its semantic content, i.e. the set of its possible denotations (as discussed in Section 6.3), and where + is used to represent event summation.

(41) 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' and 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 immediately 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.

9   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: once a linguistic semantics is defined, one can further reason on the speaker's motives for choosing one message rather than another, and for expressing it in a particular way. Although our music semantics is based on entirely different principles from linguistic semantics, it too can be expected to give rise to a pragmatics. In particular, music can be construed as being produced by a 'musical narrator' whose motives one can draw inferences about. 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 for future research to address them in greater depth.

---

[31] Pictures extracted from an animation (endlessreference), retrieved online on January 13, 2018 at https://www.youtube.com/watch?v=ZPI7_oVNB24.

9.1 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 (4): [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 the mere form of music conveys information, and is structured for this reason – as was the case in our syllable sequences;

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

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.[32] Now the highlighted F in (42)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 is itself the repetition of the preceding phrase. As a result, playing this note with an accent (louder, possibly longer) than the 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 (42)b, the accent is arguably more natural, presumably because the note is now unexpected and provides new information.[33]

(42) Modification of Ah vous dirai-je maman, with triplets[34]
      a. Simple version with triplets
      [**AV33a** http://bit.ly/2D9mds3]

      b. Modified version with with an A replacing the highlighted F in a. [**AV33b** http://bit.ly/2AVpJjM]

      c. Modified version with a D replacing the highlighted F in a. [**AV33c** http://bit.ly/2D9ZY4p]

A schematic attempt to illustrate a possible instance of Case (ii) is given in (43). Here we contrast a normal, major version of Ah vous dirai-je maman with one 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 (43)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 (43)a, the result is a bit odd, because nothing justifies highlighting this note.

---

[32] Thanks to A. Bonetto for suggesting that we consider a version with triplets.

[33] As E. Chemla and J. Katz (p.c.) note, further examples would be needed to ensure that the effect is due to a new *note* rather than to a new *contour* (since in (42)a the contour of the focused triplet is flat, just like that of its antecedent, whereas in (42)b,c the contour is not flat).

[34] The sound examples were produced as follows: Bonetto produced (42)c on an electronic piano. (42)a,b were produced from the recorded version of (42)c via manipulations (with the software GarageBand).

(43) 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. [**AV34a** http://bit.ly/2DaE9lg]

    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. [**AV34b** http://bit.ly/2EEjIdg]

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 proposed. We mention this possibility because it highlights one role of pragmatics in music, involving information structure.

9.2 Levels of intentionality

More generally, linguistic 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 is thus worth distinguishing the various levels at which intentional effects could arise. These distinctions could matter in the analysis of musical pieces.

      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 (15) 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 agents trying to achieve certain goals, as we saw in Section 3.1.

      Second, a musical piece is usually understood to be itself an intentional product: its form as well as the meaning it conveys can be attributed to an intentional agent. Let us call this agent the musical 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 performed 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 intentional levels, thereby bringing their own intentionality to the fore.[35]

9.3 Dialogues

Up to this point we have assumed that there exists only one narrator per musical piece. But once music is endowed with a semantics, a piece could also involve a *dialogue* between different narrators. This possibility might be instantiated in chamber music, with each instrument corresponding, not to an object, but to a narrator. However, detailed work would be needed to distinguish – probably on a

---

[35] As an example, consider a piece 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) [**AV35** http://bit.ly/2mBm7yR], 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 diminuendo, so that this final crescendo can't coherently be attributed to the virtual sources. Nor is it natural to think that the narrator intends this final crescendo, which contradicts the musical intention that can be inferred from the diminunendo 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.

case-by-case basis – among two interpretations. One is that each instrument is treated as a voice within our basic semantic analysis, and thus as the auditory trace of an object. This would still allow the voices to denote intentional objects and to interact in complex ways – as is the case with Heider and Simmel's animated geometric shapes, or with dancers that interact with each other intentionally without thereby *talking* to each other. An alternative is that each instrument corresponds to a narrator, and that there is genuinely a dialogue between them (here the point of comparison should be actors involved in a dialogue, rather than dancers interacting with each other). One would of course expect the dialogical interpretation to be particularly salient in opera, but in this case extra-musical cues might be so strong (due to the presence of human characters singing spoken words) as to make it hard to discern the specifically musical means that trigger this interpretation.

Still, one relatively clear instrumental case can be found in Charles Ives's Unanswered Question. Ives's Foreword describes it as follows (Ives 1908):[36]

The strings play *ppp* throughout with no change in tempo. They are to represent "The Silences of the Druids - Who Know, See and Hear Nothing." The trumpet intones "The Perennial Question of Existence", and states it in the same tone of voice each time. But the hunt for "The Invisible Answer" undertaken by the flutes and other human beings, becomes gradually more active, faster and louder through an *animando* to a *con fuoco*.

Certainly a listener who hadn't read the title or the foreword wouldn't be able to draw such specific inferences. However, our impression is that the existence of a dialogue between the trumpet and the flutes is easily perceptible by a naive listener. The trumpet alternates between the patterns in (44)a and (44)b, which are identical except for the last note (the position within the bar varies as well from one iteration to the next). The flutes reply, although not right away – the initial answer comes more than two bars after the initial question, and in later cycles the answer is heard increasingly early, and (in Ives's words) "becomes gradually more active, faster and louder through an *animando* to a *con fuoco*".

(44) Ives's Question (The Unanswered Question) [**AV37** http://bit.ly/2D5BuWH]

We believe that several factors conspire to make the dialogical interpretation of the interaction between the trumpet and the flutes very salient. Timbre certainly plays a role: wind instruments are somewhat reminiscent of the human voice, and they are culturally used to convey messages at a distance, which might help bring out the semantic interpretation of the passage. The replies don't come right away, but take some time – which is indicative of an interaction that is not directly physical, and is thus consistent with a dialogue. The question gets repeated in near-identical form six times, and each answer comes a little bit earlier than the preceding one. The melody of the question probably plays a role as well, with a long Eb that might be interpreted as carrying a special meaning or even of being focused. And the fact that the answers seem chaotic and thus unsatisfying can further explain why the question gets repeated.

While a more systematic analysis would be needed to establish whether the dialogical interpretation is indeed the salient one, and if so why, the foregoing remarks suggest that there is a natural conceptual distinction between dialogical and non-dialogical interpretations of musical pieces, and that the dialogical interpretation might indeed be favored in certain cases (there might be ambiguities in many other cases).

---

[36] In one of his *Young People's Concerts* devoted to Charles Ives, Leonard Bernstein (1967) discusses The Unanswered Question in insightful terms [**AV36** http://bit.ly/2DgUfdS] (he also adds a meta-musical reinterpretation of Ives's Unanswered Question, replacing the 'question of existence' with the question: 'Whither music?'; this is of no relevance here.)

10   Emotions

10.1   Emotional levels

The semantic content of music is often discussed in terms of emotions. These have been absent from our foregoing discussion. Do they have a natural place in our source-based analysis? We will argue that our framework has a natural place for emotions on at least four levels, corresponding to the virtual source, the listener, the musical narrator, and the musician. At the first level, the effects are squarely semantic: music may depict the emotions of some virtual sources. But we will also argue that a small modification of our framework might explain why music is particularly well suited to convey emotions. The reason is that musical patterns of tonal tension and relaxation may be easier to interpret in terms of *experienced* events, infused with emotions, than in terms of objective events. This suggests an extension of our semantic analysis: by expanding the set of possible denotations to include experienced events, more plausible interpretations of musical examples will be obtained, with a special role assigned to emotions.[37]

10.2   Types of emotional inferences

We should set aside at the outset effects that stem from the ability of sound to cause emotions irrespective of its semantics. An extremely loud sound may cause fear. Arnal et al. 2015 show that an acoustic property of human screams called 'roughness' (corresponding to amplitude modulations ranging from 30 to 150 Hz) specifically targets subcortical brain areas involved in danger processing – and of course does so irrespective of any semantics. Somewhat closer to our topic, Bonin et al. 2016 state a 'source dilemma' hypothesis according to which "uncertainty in the number, identity or location of sound objects elicits unpleasant emotions by presenting the auditory system with an incoherent percept" – and they show experimentally that subjects rate "congruent auditory scene cues as more pleasant than melodies with incongruent auditory scene cues." Here it is not so much the inferences about sources that yield emotions as the *difficulty* of identifying the sources. From a broader theoretical perspective, Huron 2006 argues that various emotions of a musical or extra-musical nature derive from general properties of expectation, i.e. of our attempts to anticipate what will come next, in music or elsewhere. But as we mentioned in Section 2.3, Huron's analysis need not depend on the existence of a music semantics. We focus the rest of this discussion on those emotion attributions that interact with our semantics.

To motivate our source-based semantics, we cited above Lerdahl's (2001) analogy between music and Heider and Simmel's (1944) abstract animations, with musical events behaving "like interacting agents that move and swerve in time and space, attracting and repelling, tensing and coming to rest". While virtual sources need not be interpreted as animate, when they are their behavior may also be indicative of emotions. As is the case more generally, inferences may be drawn on the basis both of normal auditory cognition and of the interaction between the sources and tonal pitch space (numerous tonal and non-tonal means of conveying musical emotions are surveyed in Gabrielsson and Lindström 2010, who provide a summary of experimental studies).

10.2.1   Inferences from normal auditory cognition

Inferences from normal auditory cognition have been explored in detail in the recent experimental literature, with imitations of animal signals and of human speech as primary mechanisms of inference. As mentioned above, Blumstein et al. 2012 argue that adding distortion noise (nonlinearities) in a musical piece induces in listeners an effect of "increased arousal (i.e. perceived emotional stimulation) and negative valence (i.e. perceived degree of negativity or sadness)", and they argue that such "harsh, nonlinear vocalizations" are produced by many vertebrates when alarmed, possibly because they "are produced when acoustic production systems (vocal cords and syrinxes) are overblown in stressful, dangerous situations". As was also mentioned, Bowling et al. 2010 seek to

---

[37] This section solely seeks to establish the main distinctions as they relate to music semantics; we do not do justice to the vast literature on emotions in music and in art in general (see Juslin and Sloboda 2010 for a collection of survey articles on music and emotion).

find correlates of major vs. minor intervals in excited vs. subdued speech, which might explain some of the emotional associations with these intervals.[38]

　　More generally, Juslin and Laukka 2003 propose a 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 (45) (F0 = fundamental frequency). 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.

(45)　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

　　In addition, Sievers et al. 2013 posit homologies between the mechanisms that trigger emotions in music and in the movement of a ball that can take various shapes. Specifically, they show experimentally that features that can plausibly be matched across domains (rate, jitter, i.e. regularity of rate, direction, step size, and dissonance/visual spikiness) give rise to similar emotions with music and with movement, and moreover that the finding holds across very different cultures.[39]

　　A simple example will make these points clear. 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 impression of a funeral procession is lost as the music is raised in pitch and in speed, as seen in (46). We conjecture that similar effects would be obtained with human voice, for instance, with greater speed and higher pitch (for a given voice) associated with greater animation and possibly happiness.

(46)　Mahler's Frère Jacques, measures 3-6
　　　a. Original version [**AV38a**　http://bit.ly/2r7E9O4]

　　　b. Original version, 2.5 times as fast [**AV38b** http://bit.ly/2myqJ8m]
　　　The impression of solemnity disppears.

　　　c. Original version, 2 octaves up [**AV38c** http://bit.ly/2mvQuWz]

---

[38] Bowling et al. 2012 further claim that interval size is correlated with affect in language and in music: "in both Tamil and English speech negative/subdued affect is characterized by relatively small prosodic intervals, whereas positive/excited affect is characterized by relatively large prosodic intervals"; similarly, in both Carnatic and Western music melodic intervals "are generally larger in melodies associated with positive/excited emotion, and smaller in melodies associated with negative/subdued emotion".

[39] Due to the role of dissonance in some phenomena of the natural world, it is not always clear whether certain tonal/atonal effects should be attributed to normal auditory cognition or to the interaction of the voices with tonal pitch space. It does not follow from this that the latter component can be eliminated, since the details of tonal pitch space are not given by normal auditory cognition, and may be culturally determined as well.

d. Original version, 2 octaves up, 2.5 times as fast [**AV38d** http://bit.ly/2DfpYeo]
The piece seems much happier than in the original version.

Loudness and melodic line can have powerful emotional effects as well. Let us consider a striking passage at 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 (47) is made of two ascending chromatic sequences 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 two-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 development and the intensity of Simon's discomfort: loudness and melodic height are both indicative of the strength of the discomfort.

(47) Verdi - Simon Boccanegra, Act III, Scene 3 (partial score: Simon and violins) [**AV39** http://bit.ly/2DwsA5n]
'My head is burning, I feel a dreadful fire creeping through my veins

In a non-musical domain, Aucouturier et al. 2016 showed that acoustic manipulations of a human voice can significantly affect the emotions it conveys.[40] Strikingly, one manipulation, involving an 'afraid' condition, involved a vocal version of vibrato[41]; other manipulations yielded a 'happy' or a 'sad' condition. It is likely that whatever explains these emotional effects with voices will trigger related interpretations in music. In particular, while musical vibrato needn't always produce an impression of fear, it does seem to be associated with heightened emotions – possibly because it is

---

[40] Aucouturier et al. 2016 also manipulated their subjects' voice *in real time* with similar means (e.g. addition of a vibrato). Spectacularly, they showed that subjects hearing their own manipulated voice through earphones mostly fail to detect something abnormal, but that the manipulation nonetheless *affects their own emotional state* – as if they monitored it by way of voice cues.

[41] Aucouturier et al.'s vibrato manipulation is illustrated in (i)b ("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).

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

suggestive of decreased vocal control by the source. Be that as it may, it is 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.

10.2.2 Inferences from tonal properties

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 (48) of the four (minor) realizations of the beginning of Mahler's Frère Jacques already discussed in (46).

(48) Mahler's Frère Jacques, measures 3-6 – major transposition
 a. Original tempo and height [**AV40a** http://bit.ly/2myDlfI]

 b. Original version, 2.5 times as fast [**AV40b** http://bit.ly/2mBTnGo]
 c. Original version, 2 octaves up [**AV40c** http://bit.ly/2qZRAzC]

 d. Original version, 2 octaves up, 2.5 times as fast [**AV40d** http://bit.ly/2DaLJfJ]

It is safe to assume that in each case the major version sounds happier and/or more assertive than its minor counterpart.[42]

Inferences from tonal properties also played a role in the passage quoted from Simon Boccanegra in (47). Specifically, one reason these sequences could be interpreted in terms of *discomfort* (or worse) is probably due in part to their chromatic nature. This can be seen by comparing the original, chromatic version [**AV41a** http://bit.ly/2r18ygR] with one rewritten in minor mode [**AV41b** http://bit.ly/2AX31b0] or in major mode [**AV41c** http://bit.ly/2EHP0zV]: certainly the first version is more appropriate to evoke a discomfort than the latter two.

These examples highlight the importance of specifically tonal inferences on emotions. Gabrielsson and Lindström 2010 review a rich literature that provides evidence for the traditional correlation between major mode and happiness on the one hand, and minor mode and sadness on the other. It also suggests that dissonances are interpreted in terms of unpleasantness, tension and fear, among others – which is relevant to the effect produced by the chromatic series in (47).

As a result, dissonances can trigger powerful emotional inferences – rather than in terms of a physical disequilibrium, as in Saint-Saëns's Tortoises (in (18)). An extreme example is afforded by Herrmann's music for Hichcock's *Psycho* [**AV42** http://bit.ly/2mAjZGL]; a simplified piano reduction is given in (49). 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. Various other choices contribute to the impression of mental imbalance, including the *ostinato* of the basic melodic movement, and the rhythm.

---

[42] As mentioned above, Cook 2007 and Bowling et al. 2010, 2012 seek to derive some semantic differences between minor and major chords from normal auditory cognition, and thus some of these effects might conceivably fall under the category of 'normal auditory cognition'.

(49) Herrmann's Psycho – Prelude – simple piano reduction (reduction: Hal Leonard Published, modified by A. Bonetto)[43]

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 by the feeling of anguish that it produces; much is lost in the rewritten versions.

(50) Herrmann's Psycho - reduction in (49), re-written in G minor (A. Bonetto)[44]
    a. Original reduction [**AV43a** http://bit.ly/2D2NlEK]

    b. Same as in a., re-written in G minor without dissonances [**AV43b** http://bit.ly/2EH4iFt]

    c. Same as a., closer to the original harmony [**AV43c** http://bit.ly/2mtDXmL]

10.3 External vs. internal sources: a refinement

The preceding section provided the simplest mechanism of emotion attribution within our source-based system – accounting for some instances of what is called 'perceived'/'expressed' (as opposed to 'felt') emotion in the literature.[45] But when one listens to Herrmann's music for *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 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 to be explained. Since our analysis leaves entirely open what the sources of the music are conceived to be, we can treat some of them as *experienced* sources. In other words, it makes much sense to take the objects and events that our analysis posits to be *experienced* objects and events rather than purely external ones. In this way, voices may be associated with 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.

---

[43] As Arthur Bonetto notes, this piano reduction is a compromise between the Hal Leonard version and the MIDI file we modified, though closer to Hermann's original: note values are from the first source, one half step higher (to simplify our analysis and transformations); but the sixteenth note triplets and richer chords are from the second source.

[44] In greater detail, the transformations were as follows:
 (i) From (50)a to (50)b: **Bar 1**: F# > G **Bar 2**: F# > G ; B > Bb **Bars 3-4/6-7** : F > G ; Gb > G ; B > Bb **Bar 5:** C > D; B > Bb ; Ab > G ; Eb > D.
(ii) From (50)a to (50)c: same as (i), but the boxed F > G in (i) becomes F > F# instead.

[45] See Gabrielsson 2002 for a discussion of the possible relations between perceived and felt emotion, and Evans and Schubert 2008 for relevant experimental data.

10.3.1 An example

An example from Verdi's *Simon Boccanegra* will make this 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 (47), when Simon begins to feel the effects of the poison. Even *before* he drinks from the cup, the cello theme makes clear that something momentous and disturbing is happening, as seen in (51). 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 emotions which are now reflected in the music (and possibly also the forces of destiny).

(51) Verdi's Simon Boccanegra, Act II, Scene 8 [**AV44** http://bit.ly/2FEcVlr]
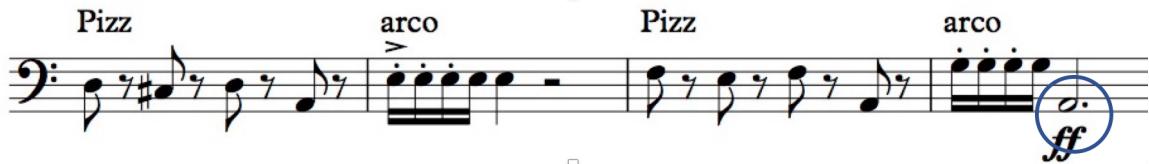
Several means conspire in the cello theme (underlined five times in (51)) to yield the impression that something momentous and disturbing is happening. The entire passage is in minor keys (arguably G minor in the first two lines and D minor in the last line). In addition, there is an alternation between slow eighth notes, with *pizzicato* timbre, 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 (52)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 La Fenice/RAI[46], Simone Piazzola as Simon drinks from the cup at exactly that point [**AV44** http://bit.ly/2FEcVlr].)

(52) a. The last line of (51) is written with a chromatic ascent and a tritone interval (boxed), yielding a feeling of tension and anguish. [**AV45a** http://bit.ly/2AYlh3J]

b. Rewriting a. in D minor (without chromatic excursions) removes much of the feeling of tension and

---

[46]Simon Boccanegra, Teatro La Fenice 2014-2015, conductor Myung-Whun Chung, RAI, with Simone Piazzola as Simon.

anguish (re-written by A. Bonetto). [**AV45b** http://bit.ly/2FCxXAQ]

10.3.2   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 (25), 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 was seen in (25)b. We could now add a further possibility, namely that O's being in e_i causes a less stable emotion than O's being in e_k. The modified Harmonic stability condition, stated in (53), is disjunctive, a property it shares with our old Loudness condition, seen in (25)a.

(53)   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.

Let us add that we were forced to stipulate certain properties of the 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, in particular to determine what a 'stable' emotion is – a non-trivial requirement.

This brief discussion of the ways in which our semantics could make provisions for experienced events is only a proof of concept. But it suggests that that there are at least two general ways in which a source-based semantics can incorporate the role of musical emotions: by way of emotions attributed to the sources when these are construed as animate; and by way of an extension of the framework in which some or all of the denoted events are experienced rather than purely external events.

11   Conclusions

11.1   Theoretical conclusions

If our proposal is on the right track, music has a semantics, but one that is closer to picture semantics than to logical semantics. We treated music cognition as being continuous with normal auditory cognition, and in both cases we took the semantic content derived from an auditory percept to be closely connected to 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, and tonal properties of the music. This made it possible to sketch a truth-conditional semantics for music: a musical 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 world events they are supposed to denote. This guaranteed in particular that music semantics is appropriately abstract: in general, there is no requirement that the denoted events should be sound-producing.

We outlined several consequences that could 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. 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. (Further extensions and questions are discussed in Appendix IV.)

11.2 *Methodological conclusions*

Although we based our theoretical discussion on informal introspective judgments (which should be subjected to experimental methods in the future), we made frequent use of 'minimal pairs' to display semantic effects – a standard approach in experimental music psychology, but possibly one that should be used more systematically when studying the effects of 'real' music.

In order to *explain* semantic effects, methods differ depending on whether they have 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 relevant properties of tonal pitch space. Importantly, the inferences that one might need to test are quite abstract in nature, and thus 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.

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-cultural investigation of music semantics should prove illuminating.

Appendix I. Varieties of internal semantics

This Appendix discusses in greater detail the notion of an 'internal semantics' for music, briefly mentioned in Section 2.2.

Before we say a word about the 'internal' semantics in music, we consider how such a semantics can be constructed for a system as simple as the *la li lu* example of the Section 2.1. The key is that a syntactic system that has no semantics relating it to the external world can still be *endowed* with a semantics that pertains to the *form* of the expressions themselves. In (54), we have done so for the context-free grammar defined in the main text 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.[47] Some simple examples are given in (55).

(54) a. Lexical semantics:
　　　$[[la]] = la$
　　　$[[lu]] = lu$
　　　$[[li]] = li$

　　　b. Compositional semantics
　　　Notation: ˆ is used to represent concatenation of expressions; for strings *s* and *s'*, *s*ˆ_ˆ*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,
　　　$[[[_L w\ w']]] = [[w]]ˆ[[w']]$
　　　$[[l\ s]] = [[l]]ˆ\_ˆ[[s]]$

(55) Examples
　　　$[[[_L la\ lu]]] = [[la]]ˆ[[lu]] = laˆlu$
　　　$[[[_L la\ lu]\ [_L la\ li]]] = [[[_L la\ lu]]]ˆ\_ˆ[[[_L la\ li]]] = laˆluˆ\_ˆlaˆli$

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 syllable sequences. For example, in (56) we define 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*ˆ*u*, i.e. the concatenation of the vowels *i* and *u*.

(56) 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,
　　　$[[l\ s]] = [[l]]ˆ[[s]]$

(57) Examples
　　　$[[[_L la\ lu]]] = u$
　　　$[[[_L la\ lu]\ [_L la\ li]]] = [[[_L la\ lu]]]ˆ[[[_L la\ li]]] = uˆi$

We can think of this semantics as associating with some strings a 'vocalic path' that tracks the sequence of some particularly important phonemes 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 (54), there are prominent examples of music semantics that develop more sophisticated versions of (56). 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 (59). 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

---

[47] Languages in which objects name themselves are called 'Lagadonian languages' in the philosophical literature (e.g. Lewis 1986).

syntax goes hand in hand with a semantic step. And the semantics encodes movements in tonal pitch space.

A minimal example is given in (58), which provides the semantics of a sequence V^7-I within a tonal pitch space whose structure is displayed in (59). 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, written as: λx. leftonto(x). Taking the location <0, 0> as an argument, the result is: leftonto(<0, 0>). Assuming the tonic (i.e. <0, 0>) is a C (circled in (59)), this would correspond to a movement from a G (also circled in (59)) to that C.

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

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

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

E	B	F♯	C♯	G♯	D♯	A♯	E♯	B♯
C	G	D	A	ⓔ	B	F♯	C♯	G♯
A♭	E♭	B♭	F	Ⓒ	Ⓖ	D	A	E
F♭	C♭	G♭	D♭	A♭	E♭	B♭	F	C
D♭♭	A♭♭	E♭♭	B♭♭	F♭	C♭	G♭	D♭	A♭

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.

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 developed in this piece.

Appendix II. Music semantics vs. logical semantics

This Appendix compares in greater detail music semantics with the simple logical semantics sketched in Section 7.1.

In order to bring the comparison into sharper focus, 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 logic is defined for a language made solely of propositional letters, and conjunctions. Since the only way to combine propositional letters is by way of conjunction, we don't need an explicit conjunction sign and thus we will solely investigate sentences of the form: p_i, $p_i p_k$, $p_i p_k p_r$, etc., as is specified by the syntax in (60)a. Each propositional letter is taken to hold true of events[48], and concatenation is interpreted as conjunction, as shown by the semantics in (60)b.

(60) A purely conjunctive logic
      a. Syntax
      (i) Atomic propositions: for every i ≥ 0, p_i is an atomic proposition.
      (ii) If p_i is an atomic proposition and F is a proposition, $p_i F$ is a proposition.

      b. Semantics
      Let I be a function such that for every i ≥ 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 atomic 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.

Let us immediately illustrate with the examples in (61). If p_2 holds true of events e, e' and e", while p_3 holds true of events e' and e", the (implicit) conjunction $p_2 p_3$ holds true of e' and e", as does $p_3 p_2$. If in addition p_1 holds true of e and e', $p_1 p_2 p_3$ holds true of e'.

(61) 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'.

We note that our rules are designed in such a way that a string is always semantically analyzed from beginning to end: a string $p_1 p_2 p_2$ is analyzed by the semantics (in (60)b(ii)) as having the structure *[p₁ [p₂ p₃]]* . Nothing deep hinges on this: given our conjunctive semantics, whether a string $p_1 p_2 p_3$ is analyzed as *[p₁ [p₂ p₃]]* (as we do) or as *[[p₁ p₂] p₃]* won't affect the truth conditions, since the end result will just be the conjunction of p_1, p_2 and p_3.

One could think of $p_1 p_2 p_3$ as a series of musical events, which may be true of some events such as e, e', e", etc. But as noted in the text, the similarities with our music semantics end there: the conjunctive logic above has no counterparts of our preservation principles (Time, Loudness, Harmonic stability); and in that logic, $p_1 p_2$ denotes events that satisfy both p_1 and p_2 (order irrelevant), rather than pairs of events $<e_1, e_2>$ that satisfy p_1 and p_2 (in that order), as in our music semantics.

---

[48] In more standard systems, propositional letters are true at certain possible worlds. Events are typically thought to be more fined-grained than worlds because distinct events co-exist in the same possible world. An event-rather than world-based semantics facilitates the comparison with our music semantics, which builds on the idea that sequences of musical notes or chords can be true of sequences of extra-musical events (treated as possibilia).

Appendix III. Complements on the syntax/semantics interface

This Appendix provides some complements to the discussion of the syntax/semantics interface in Section 8. Part A gives details about exceptions to tree structure in Lerdahl and Jackendoff's analysis, as alluded to in Section 8.3.1. Part B formulates some questions raised for the present analysis by Lerdahl and Jackendoff's prolongational reductions.

A. Exceptions to tree structure in Lerdahl and Jackendoff's analysis of grouping
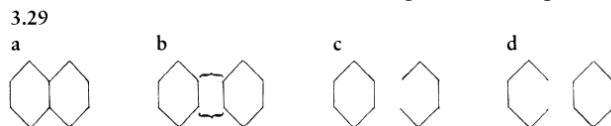
We argued in Section 8.3.1 that a mereology-based reconstruction of musical structure leads one to expect that two musical groups could have a part in common (as in (36)e in the main text) in two types of cases: when the denoted events are best analyzed as having a part in common (overlap); and when two distinct events share the same auditory trace (occlusion).

Lerdahl and Jackendoff 1983 emphasize that such cases do 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 by appealing to analogous cases in visual perception. The exceptions they list are of the two types we expect: in case of 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 (62)a, which is preferably analyzed as (62)b rather than as (62)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.

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

Cases of overlap are probably pervasive in event decomposition as well. A person's walk is a succession of cycles in which a foot touches ground, goes up, and touches ground again. Each subevent in which the foot touches ground is both the completion of a cycle and the beginning of the next one – an event counterpart of the object perception case in (62).

Lerdahl and Jackendoff 1983 cite the very beginning of Mozart's K. 279 sonata as an example of auditory overlap, as seen in (63). The I chord at the beginning of bar 3 seems both to conclude the first group and to 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 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 (62)).
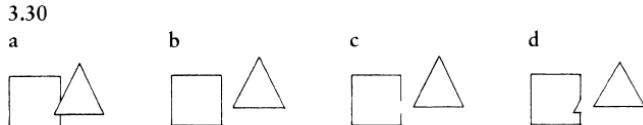
(63) An example of overlap: the beginning of Mozart's K. 279 sonata (Lerdahl and Jackendoff 1983 p. 56) [AV46 http://bit.ly/2D9Xioe]

B. 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.[49] 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 (39) in the main text. 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 ([(39)] 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)

(66) Prolongation of the initial I chord at the beginning of of Mozart's K. 331 piano sonata (Lerdahl and Jackendoff 1983) [AV48 http://bit.ly/2DamRom]

In the time-span structure in (39), 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 it corresponds to the end of a tensing and relaxing motion that started 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.

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 between 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 prolongational structures.

Second, if prolongational reductions trace the 'harmonic path' followed by virtual sources in tonal pitch space, could one also investigate other types of paths, such as those defined by the melodic line or even the evolution of loudness? A key insight of Lerdahl and Jackendoff's analysis of

---

[49]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 share properties with) syntactic structure. They further suggest that prolongational reductions need not be taken to be derivative from time-span reductions, as argued by Lerdahl and Jackendoff 1983 and Lerdahl 2001.

prolongational structure is that two harmonic events that are not linearly contiguous may still have a direct structural connection, as is the case of the two highlighted I chords in (66). But in the classic analysis of Schenker, the melodic line (and in particular a gradually descending melodic line called the 'Urlinie') has related properties: certain melodic elements can be ignored when tracing the general melodic line of a piece because they are structurally subordinate (Forte 1959, Pankhurst 2008). Could one develop a general theory in which Lerdahl and Jackendoff's prologongational reductions and Schenker's Urlinie are part of a broader typology? This and other questions pertaining to the interaction between prolongational reductions and music semantics must be left for future research.

Appendix IV. Extensions and further questions

This Appendix sketches some possible extensions of our analysis, and lays out further questions for future research.

❑ *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 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 picture in (67).

(67) Coarse-grained pictorial representation of 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', and we saw striking instances of this mechanism in Saint-Saëns's *Carnival*.
(ii) Second, when analyzing a piece, one may decide to interpret each and every note as corresponding to a 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 discussed in (20) in the main text, one may answer in the positive 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.

❑ *Interpreting a piece*

If our analysis is on the right track, a musician interpreting a piece may make musical decisions that further specify 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 may 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 must make semantic decisions that are left open by the score. Even after these decisions are made, there will be a plurality of situations that the music is compatible with (true of), but the musician's interpretation will usually reduce the set of situations that are compatible with the score.

❑ *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 offer a music aesthetics. Still, one might hope that some aesthetic considerations might in the end build on insights gained from music semantics. But nothing at all in the present enterprise suggests that the aesthetic effects of music (or for that matter its psychological effects) *reduce* to its semantics. This would be as absurd as claiming that the aesthetic or psychological effects of poetry are exhausted by its semantics.

❑ *Semantic effects beyond music*

The key idea of our semantic analysis of music is that denotational inferences may be drawn both from normal 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 [AV49 http://bit.ly/2DzS6He]: 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 represent 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 correspondence between the auditory and the visual domain can be found.[50]

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 combined (do they give rise to a single semantic representation?). For relevant work, see Charnavel 2016, Napoli and Kraus, to appear, and Patel-Grosz et al. 2017.

❑ *Further questions*

Two types of further questions could be asked. One pertains to the nature of semantic inferences in music. The other concerns the nature of the theory we have proposed.

---

[50] It would be interesting to explore in the future a closer correspondence between the musical and the visual domains, in particular by seeking visual counterparts of the various cues that trigger inferences about the sources, be they drawn from normal auditory cognition or from harmonic considerations. In this connection, simple systems of music visualization give rise to interesting abstract animations, 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 harmony. Harmonic coloring [http://www.musanim.com/mam/circle.html] has thus been used to provide an animated rendition of Stravinsky's *Rite of Spring* [AV50 http://bit.ly/2mGzeyK; http://bit.ly/2FEqYar]. But it is clear that even harmonic coloring only yields a crude (and not necessarily intuitive) encoding of the complex harmonic relations among notes and chords.

One could ask whether semantic inferences in music are always conscious.[51] This need not be the case: we had to create minimal pairs and to ask appropriately abstract questions about the virtual sources in order to bring these inferences to consciousness. Certainly this is a matter of degree: some semantic effects are more subtle than others. It might be that, quite generally, inferences that are abstract and hard to put in words tend to be less conscious than more concrete and easily articulable ones (one could explore this question in other cognitive domains, such as visual cognition or the recognition of tastes and odors). Be that as it may, we believe that even when semantic inferences are unconscious, they are crucial to explain some of the psychological effects of music, as well as interpretive choices made by performers.

Turning to the nature of the theory we have proposed, one could ask whether we have not overly stretched the extension of the term 'semantics'.[52] Our source-based semantics is in part based on our ability to draw inferences on the *causal* sources of a sound. One could object that the term 'semantics' is properly used only if it pertains to an *intention* to mean something, perhaps along the lines of Grice's notion of 'utterer's meaning' (Grice 1957). Without taking a position on this issue, we note that our final account *does* have a place for the equivalent of a kind of utterer's meaning, since the existence of a music semantics makes it possible to reconstruct the intentions of a musical narrator that attempts to convey a certain (highly abstract) message. The situation is in this respect no different from that of a narrator expressing herself in gestures or by way of drawings or visual animations.

---

[51] Thanks to B. Spector (p.c.) for raising this question.
[52] Thanks to J. MacFarlane (p.c.) for raising this question.

Audiovisual examples

The audiovisual examples can be downloaded at the following URL: http://bit.ly/2DBhNH6

Credits for audiovisual examples

AV00 Kominsky, J.F., Strickland, B., Wertz, A.E., Elsner, C., Wynn, K., & Keil, F.C.: 2017, Categories and constraints in causal perception. *Psychological Science* 28,11: 1649-1662. Video of the 'violation' condition (of laws of elastic collision). (Thanks to B. Strickland for providing this video.)

AV01 Musicnet materials, retrieved online on January 7, 2018 at https://www.youtube.com/watch?v=UH3IULiXo24.

AV02 Il con, Stanley Kubrick, *2001: A Space Odyssey*. Retrieved online on January 7, 2018 at https://www.youtube.com/watch?v=e-QFj59PON4.

AV05, AV06, AV07, AV08, AV09 Musicanth materials, retrieved online on January 9, 2018 at https://www.youtube.com/watch?v=5LOFhsksAYw. Pianists: Vivian Troon, Roderick Elms. Conductor: Andrea Licata Royal Philharmonic Orchestra.

AV13 Video cited in: Honing, H.: 2003, The final ritard: On music, motion, and kinematic models. *Computer Music Journal*, 27(3), 66-72.

AV17 Kurt Masur: Leipzig Gewandhaus Orchestra.

AV18 Musicanth materials, retrieved online on January 9, 2018 at https://www.youtube.com/watch?v=5LOFhsksAYw. Pianists: Vivian Troon, Roderick Elms. Conductor: Andrea Licata Royal Philharmonic Orchestra.

AV20 Mozart, Don Giovanni, Metropolitan Opera, 1990, directed by Brian Large, stage production Franco Zeffirelli, conductor James Levine, with Kurt Moll as the Commendatore.

AV23 Tchaikovsky Ouverture solennelle "1812," Op.49, by Berliner Philharmoniker. Conductor Claudio Abbado. Retrieved online on January 9, 2018 at https://www.youtube.com/watch?v=lDIVz9r3q3s.

AV24 Puccini, Madama Butterfly. Asheville Lyric Opera. Pinkerton - Brian Cheney; Sharpless - Mark Owen Davis. Retrieved online on January 9, 2018 at https://www.youtube.com/watch?v=YgLQ3p6hSOI.

AV25, AV28 Musicanth materials, retrieved online on January 9, 2018 at https://www.youtube.com/watch?v=5LOFhsksAYw. Pianists: Vivian Troon, Roderick Elms. Conductor: Andrea Licata. Royal Philharmonic Orchestra.

AV31 Mary Ellen Bute (visuals) and Edwin Gerschefski (piano accompaniment), 1940, Tarantella (excerpt starting at 1:09). Retrieved online on January 13, 2018 at https://www.dailymotion.com/video/xgd0wn.

AV32 Daniel Barenboim, Mozart: Complete Piano Sonatas and Variations, Piano Sonata No. 1 in C, K.279: I. Allegro.

AV36 Leonard Bernstein: Young People's Concerts Vol. 2. Charles Ives American Pioneer (excerpt starting at 44:33). Retrieved online on January 14, 2018 at https://www.youtube.com/watch?v=tsbaSwhtx9E.

AV37 Charles Ives, The Unanswered Question. James Sinclair, Conductor. Northern Sinfonia (excerpt starting at 1:00). Retrieved online on January 14, 2018 at https://www.youtube.com/watch?v=kkaOz48cq2g.

AV38 Verdi, Simon Boccanegra, Teatro La Fenice 2014-2015, conductor Myung-Whun Chung, RAI, with Simone Piazzola as Simon.

AV42 *Psycho*, 1960. Film directed and produced by Alfred Hitchcock, and written by Joseph Stefano. Music by Bernard Herrmann.

AV44 Verdi, Simon Boccanegra, Teatro La Fenice 2014-2015, conductor Myung-Whun Chung, RAI, with Simone Piazzola as Simon.

AV46 Daniel Barenboim, Mozart: Complete Piano Sonatas and Variations, Piano Sonata No. 1 in C, K.279: I. Allegro.

AV47 Haydn: Symphony #104 In D, H 1/104, "London" - 1. Adagio - Allegro. Adam Fischer: Austro-Hungarian Haydn Orchestra (starting at 2:29).

AV48 Margarete Babinsky. Piano Sonata in A, K.331: I. Andante Grazioso. Mozart Piano Sonatas.

AV49 Todd Mostak, Twitter Heatmapper. Retrieved online on January 14, 2018 from https://www.youtube.com/watch?v=4_v2EZGiA7w.

AV50 Musicanim, graphical score generated for Stravinsky's The Rite of Spring, on the basis of Jay Bacal's synthetic MIDI recording. Retrieved online on January 14, 2018 from https://www.youtube.com/watch?v=5IXMpUhuBMs.

References

Arnal, LH; Flinker, A; Kleinschmidt, A; Giraud, AL; Poeppel, D: 2015, Human screams occupy a privileged niche in the communication soundscape. *Current Biology* 25 (15), 2051-2056

Atkin, Albert: 2013, Peirce's Theory of Signs. *The Stanford Encyclopedia of Philosophy* (Summer 2013 Edition), Edward N. Zalta (ed.), URL = <https://plato.stanford.edu/archives/sum2013/entries/peirce-semiotics/>.

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.

Berg Larsen, Joakim: 2017, *Conceptions of Meaning in Music. On the Possibility of Meaning in Absolute Music*. Master's Thesis in Philosophy, Arctic University of Norway. Retrieved online at https://munin.uit.no/bitstream/handle/10037/11806/thesis.pdf on January 6, 2018.

Bernstein, Leonard: 1967, Charles Ives: American Pioneer. Young People's Concerts. Television series, February 23, 1967.

Blumstein, Daniel T., Bryant, Gregory A. and Kaye, Peter: 2012, The sound of arousal in music is context-dependent. *Biol. Lett.* 8, 744-747

Bonin, T.L., Trainor, J.L., Belyk, M. & Andrews, P.: 2016, The source dilemma hypothesis: Perceptual uncertainty contributes to musical emotion. *Cognition* 154:174-181.

Bowling DL, Gill K, Choi J, Prinz J, Purves D: 2010, Major and minor music compared to excited and subdued speech. *Journal of the Acoustical Society of America* 127: 491–503.

Bowling DL, Sundararajan J, Han S, Purves D: 2012, Expression of Emotion in Eastern and Western Music Mirrors Vocalization. *PLoS ONE* 7(3): e31942. doi:10.1371/journal.pone.0031942

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.

Clarke, Eric: 2001, Meaning and the specification of motion in music. *Musicae Scientiae* 5: 213–34.

Cohn, N.; Jackendoff, R.; Holcomb, P. J.; Kuperberg, G. R.: 2014, The grammar of visual narrative: Neural evidence for constituent structure in sequential image comprehension . *Neuropsychologia*, 64: 63 ─ 70.

Cook, Norman D.: 2007, The Sound Symbolism of Major and Minor Harmonies, *Music Perception*, 24: 315–19.

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.

Davidson, Donald: 1967, The logical form of action sentences. In N. Rescher (ed.) *The logic of decision and action*. Pittsburgh: University of Pittsburgh Press. 81-94.

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.

Evans, P., & Schubert, E.: 2008, Relationships between expressed and felt emotions in music. *Musicae Scientiae*, 12, 75-99.

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.

Gabrielsson, Alf and Lindström, Eric: 2010, The role of structure in the musical expression of emotions. In: *Handbook of Music and Emotion: Theory, Research, and Applications*, eds Juslin P. N., Sloboda J. A., Oxford: Oxford University Press, 367–400

Gabrielsson, A.: 2002, Emotion perceived and emotion felt: Same or different? *Musicae Scientiae*, Special Issue, 123-147

Godoy, R. I. and Leman, M. (Eds.): 2010, *Musical gestures: Sound, movement, and meaning*. Routledge.

Granroth-Wilding, Mark and Steedman, Mark: 2014, A robust parser-interpreter for jazz chord sequences. *Journal of New Music Research* 43 (4), 355-374.

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

Grice, Paul: 1957, Meaning. *The Philosophical Review* 66: 377–88.

Heider, F., and Simmel, M.: 1944, An experimental study of apparent behavior. *American Journal of Psychology*, 57, 243-259.

Heim, Irene and Kratzer, Angelika: 1998, *Semantics in Generative Grammar*. Blackwell.

Honing, H.: 2003, The final ritard: On music, motion, and kinematic models. Computer Music Journal, 27(3), 66-72.

Huron, David: 2006, *Sweet Anticipation: Music and the Psychology of Expectation*. Cambridge, MA: MIT Press.

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.

Huron, David: 2016, *Voice Leading: The Science Behind a Musical Art*. MIT Press.

Ilie, G. and Thompson,W. F.: 2006, A comparison of acoustic cues in music and speech for three dimensions of affect. *Music Perception*, 23, 319–29.

Ives, Charles: 1908, Foreword to *The Unanswered Question*. Available online at http://www.dmu.uem.br/aulas/historia3/Aula11_Ives/The_Unanswered_Question.pdf

Jackendoff, Ray: 1982, *Semantics and Cognition*. MIT Press.

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.

Juslin P. N., Sloboda J. A. (eds): 2010, *Handbook of Music and Emotion: Theory, Research, and Applications*. Oxford: Oxford University Press

Kivy, Peter: 1990, *Music Alone*. Ithaca: Cornell University Press.

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

Koelsch, S.: 2012, 'Musical Semantics'. Chapter from *Brain and Music*, Wiley-Blackwell.

Koelsch, S., Kasper, E., Sammler, D., Schulze, K., Gunter, T., & Friederici, A. D.: 2004, Music, language and meaning: brain signatures of semantic processing. *Nature neuroscience*, 7(3), 302-307.

Kominsky, J.F., Strickland, B., Wertz, A.E., Elsner, C., Wynn, K., & Keil, F.C.: 2017, Categories and constraints in causal perception. *Psychological Science* 28,11: 1649-1662

Kracht, Marcus: 2003, *The Mathematics of Language (Studies in Generative Grammar, 63)*, Mouton de Gruyter.

Larson, Richard: 1995, Semantics. n D. Osherson, L. Gleitman and M. Liberman (eds), An Invitation to Cognitive Science, Vol. I: Language, Cambridge, Mass.: MIT Press.

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.

Lerdahl, Fred and Carol L. Krumhansl: 2007, Modeling Tonal Tension. *Music Perception*, 24.4, 329–366

Lewis, David K.: 1970, General semantics. *Synthese*, 22(1/2), 18–67. Repr. 1983, in Philosophical Papers, Vol. I: 189–229. Oxford University Press.

Lewis, David 1979. Attitudes de dicto and de se. *Philosophical Review* 88 (4): 513-543

Lewis, David K: 1986, *On the Plurality of Worlds*. Blackwell, Oxford.

Longuet-Higgins, H. C.: 1962a, Letter to a musical friend. *The Music Review*, 23:244–248.

Longuet-Higgins, H. C.: 1962b, Second letter to a musical friend. *The Music Review*, 23:271–280.

McCawley, James. 1982. Parentheticals and discontinuous constituent structure. *Linguistic Inquiry* 13, 1, 91–106.

Maus, Fred Everett: 1988, Music as Drama. *Music Theory Spectrum*, Vol. 10, 10th Anniversary Issue (Spring, 1988), pp. 56-73

McDermott JH, Lehr AJ, Oxenham AJ: 2010, Individual differences reveal the basis of consonance. *Current Biology* 20: 1035–1041.

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

Monahan, Seth: 2013, Action and Agency Revisited. *Journal of Music Theory* 57:2

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

Nudds, Matthew: 2007, Auditory Perception and Sounds. Manuscript.

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.

Pankhurst, Tom: 2008, *Schenkerguide. A Brief Handbook and Website for Schenkerian Analysis*. Routledge.

Patel-Grosz, Pritty; Grosz, Patrick G.; Kelkar, ,Tejaswinee; Jensenius, Alexander Refsum : 2017, Exploring the Semantics of Dance. Slides of a talk given at the Harvard Language & Cognition lab (LangCog), 14 Feb 2017.

Peirce, Charles S.: 1868, On a New List of Categories. *Proceedings of the American Academy of Arts and Sciences* 7, 287-298.

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

Rohrmeier, Martin: 2011, Towards a generative syntax of tonal harmony. *Journal of Mathematics and Music* 5

(1), 35–53.

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

Rosner, B. S., & Narmour, E.: 1992, Harmonic Closure: Music Theory and Perception. *Music Perception* 9 (4), 383-411. http://dx.doi.org/10.2307/40285561

Saslaw, Janna: 1996, Forces, Containers, and Paths: The Role of Body-Derived Image Schemas in the Conceptualization of Music. *Journal of Music Theory* 40, no. 2: 217–43

Schlenker, Philippe: 2010, Semantics. In *Linguistics Encyclopedia*, ed. K. Malmkjaer, 3rd edition, Routledge, pp. 462-477.

Schlenker, Philippe: 2011, Indexicality and De Se Reports. In *Semantics,* edited by von Heusinger, Maienborn and Portner, Volume 2, Article 61, pp. 1561-1604, Mouton de Gruyter.

Schlenker, Philippe: 2017, Outline of Music Semantics. *Music Perception: An Interdisciplinary Journal* 35, 1: 3-37 DOI: https://doi.org/10.1525/mp.2017.35.1.3

Schlenker, Philippe: to appear, Iconic Pragmatics. *Natural Language & Linguistic Theory*.

Schlenker, Philippe, Lamberton, Jonathan and Santoro, Mirko: 2013, Iconic Variables. *Linguistics & Philosophy* 36, 2, 91-149.

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

Sportiche, Dominique; Koopman, Hilda; Stabler, Ed: 2013, *An Introduction to Syntactic Analysis and Theory*. Wiley-Blackwell.

Thompson, W. F., & Cuddy, L. L.: 1992, Perceived key movement in four-voice harmony and single voices. *Music Perception*, 9(4), 427-438.

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.