Outline of Music Semantics*

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

May 21, 2016 - Draft 1.1
(minor corrections)

[This is a summary of 'Prolegomena to Music Semantics' [http://ling.auf.net/lingbuzz/002925], which discusses several issues that are omitted here – notably: the role of emotions in music semantics, music pragmatics, and the radical differences between music semantics and linguistic semantics.]

Music consultant: Arthur Bonetto

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

Abstract. We provide the outline of a semantics for music. We take music cognition to be continuous with normal auditory cognition, and thus to deliver inferences about 'virtual sources' of the music (as in Bregman's Auditory Scene Analysis). As a result, sound parameters that trigger inferences about sound sources in normal auditory cognition produce related ones in music – as is the case when decreasing loudness signals the end of a piece because the source is gradually losing energy, or moving away. But what is special about music is that it also triggers inferences on the basis of the movement of virtual sources in tonal pitch space, which has points of stability (e.g. a tonic chord), points of instability (e.g. dissonant chords), and relations of attractions among them (e.g. a dissonant chord tends to be resolved). In this way, gradual movement towards a point of tonal stability, as in a cadence, may also serve to signal the end of a piece, but on the basis of tonal information. The challenge is thus to develop a framework that aggregates inferences from normal auditory cognition and tonal inferences. We sketch such a framework in a highly simplified case, by arguing that a source 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*. Thus we require that inferences triggered by loudness on the relative levels of energy or proximity among events should be preserved, and similarly for tonal inferences pertaining to the relative stability of events. This yields a 'bare bones' version of a music semantics, as well as a definition of 'musical truth'. We then argue that this framework can help re-visit some aspects of musical syntax. Specifically, we take (Lerdahl and Jackendoff's) *grouping structure* to reflect the mereology ('partology') of events that are abstractly represented in the music – hence the importance of Gestalt principles of perception in defining musical groups.

---

* The present manuscript presents the 'bare bones' of an account sketched in greater detail in Schlenker 2016. Special thanks Arthur Bonetto served as a regular and very insightful music consultant for these investigations; virtually all musical examples were discussed with him, and he played a key role in the construction of all minimal pairs, especially when a piece had to be rewritten with special harmonic constraints. However he bears no responsibility for theoretical claims – and possible errors – contained in this piece.
For helpful conversations, many thanks to John Bailyn, Karol Beffa, Arthur Bonetto, Laurent Bonnasse-Gahot, Emmanuel Chemla, Didier Demolin, Paul Egré, John Halle, Ray Jackendoff, Jonah Katz, Fred Lerdahl, Salvador Mascarenhas, Rob Pasternak, Claire Pelofi, Martin Rohrmeier, Benjamin Spector, Morton Subotnick, Francis Wolff, as well as audiences at New York University and SUNY Long Island. I learned much from initial conversations with Morton Subotnick before this project was conceived. Jonah Katz's presence in Paris a few years ago, and continued conversations with him, were very helpful. I have also benefited from Emmanuel Chemla's insightful comments on many aspects of this project. None of these colleagues is responsible for any errors made here.
The research leading to these results received funding from the European Research Council under the European Union's Seventh Framework Programme (FP/2007-2013) / ERC Grant Agreement N°324115–FRONTSEM (PI: Schlenker). Research was conducted at Institut d'Etudes Cognitives, Ecole Normale Supérieure - PSL Research University. Institut d'Etudes Cognitives is supported by grants ANR-10-LABX-0087 IEC et ANR-10-IDEX-0001-02 PSL*.

1 Introduction

Can there be a music semantics, construed as *a set of principles that determine the informational content of a musical piece about a music-external reality*? Despite much discussion of the 'meaning of music', this question remains opaque. By contrast, the formal study of musical syntax made considerable progress in the last 40 years (see for instance Jackendoff and Lerdahl 1983, Lerdahl 2001 for classical music, and Granroth-Wilding and Steedeman 2014 for jazz), leading to precise and empirically predictive theories of possible musical sequences. Can formal and empirical progress be made on the semantic front as well?

In this short paper, we sketch a positive (and programmatic) answer. Our guiding intuition is that *the meaning of a musical piece is given by the inferences that one can draw about its 'virtual sources',* which in salient cases can be identified with the 'voices' of classical music theory. Our analysis is in two steps.

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

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

To see an example of an inference from *normal auditory cognition*, consider loudness. One common way to signal the end of a piece is to gradually decrease the loudness. While this device may be taken to be conventional, it is plausible that it is derived from normal auditory audition: a source that produces increasingly softer sound may be losing energy, or moving away (we will see below that both types of inference can be triggered). Turning to *tonal inferences*, 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 in a tonic position is in the most stable tonal position, hence not 'attracted' to other parts of tonal pitch space. In quite a few cases, these two types of inference conspire to mark the end of a piece. But as we will argue, numerous other semantic effects can be produced as well.

The challenge is thus to develop a framework that aggregates inferences from normal auditory cognition and tonal inferences. We sketch such a framework in a highly simplified case, by arguing that a source 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*. Thus we require that inferences triggered by loudness on the relative levels of energy or proximity among events should be preserved, and similarly for tonal inferences pertaining to the relative stability of events. This yields a 'bare bones' version of a music semantics, as well as a definition of 'musical truth'. It is obtained from entirely different means from truth in language: musical inferences are draw by treating music as the 'auditory trace' of some abstract sources, and not by a compositional procedure, as in human language.

We will then argue that this framework can help re-visit some aspects of musical syntax, along the lines of Lerdahl and Jackendoff 1983. Specifically, we will take their *grouping structure* to reflect the mereology ('partology') of events that are abstractly represented in the music – hence the importance of Gestalt principles of perception (rather than of a generative syntax) in defining musical groups. In other words, we take musical grouping to originate in an attempt to reconstruct the structure of events undergone by the virtual source. In many cases, mereological relations yield a tree-like structure for groups, but as was already discussed in Lerdahl and Jackendoff 1983, there are some exceptions; we will argue that they follow from the mereological interpretation. We will further speculate that the asymmetric, 'headed' nature of musical groups (corresponding to Lerdahl and Jackendoff's 'time-span reductions') might reflect a more general tendency to analyze sub-events as being more or less important for events they are part of.

To situate our enterprise, our framework seeks to integrate two intuitions that were developed in earlier analyses.

• In Bregman's application of Auditory Scene Analysis to music, the listener analyzes the music as a kind of 'chimeric sound' which 'does not belong to any single environmental object' (Bregman 1994 chapter 5). As Bregman puts it, 'in order to create a virtual source, music manipulates the factors that control the formation of sequential and simultaneous streams'. Importantly, 'the virtual source in music plays the same perceptual role as our perception of a real source does in natural environments', which allows the listener to draw inferences about the virtual sources of the music (in Bregman's terms, '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').

• The other antecedent idea is that the semantic content of a musical piece is a kind of 'journey through tonal pitch space'. Lerdahl 2001 thus analyzes 'musical narrativity' connection with a linguistic theory (Jackendoff 1982) in which 'verbs and prepositions specify places in relation to starting, intermediate, and terminating objects'. For him, music is equally 'implicated in space and 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, Granroth-Wilding and Steedman 2014 provide an explicit semantics for jazz sequences in terms of motion in tonal pitch space.

It is essential for us that that these two ideas should be combined within a single framework. An analysis based on Auditory Scene Analysis alone might go far in identifying the virtual sources and explaining some inferences they trigger on the basis of normal auditory cognition, but it would fail to account for the further inferences that one draws by observing the movement of the voices in tonal pitch space – for instance the fact that 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 (as in the case of decreasing volume), and more generally it would miss the fact that the sources can be construed as real world objects – an essential condition to obtain a *bona fide* semantics.

The rest of this article is organized as follows. In Section 2, we motivate our project by discussing an example of a semantic effect in music, based on inferences from normal auditory cognition and tonal properties; this also allows us to explain the import of the method of 'minimal pairs' to help bring out subtle semantic effects. In Section 3, we list some types of inference drawn from normal auditory cognition, and we do the same thing for inferences from tonal properties in Section 4. In Section 5, we explain how both types of inference can be aggregated in a highly simplified theory of 'musical truth'. We then argue in Section 6 that the our semantic approach can account for some aspects of musical syntax, including cases in which tree structures turn out to be inadequate. Conclusions and questions for future research are listed in Section 7.

2 Motivating Music Semantics

2.1 The Null Theory: no semantics or an 'internal' semantics

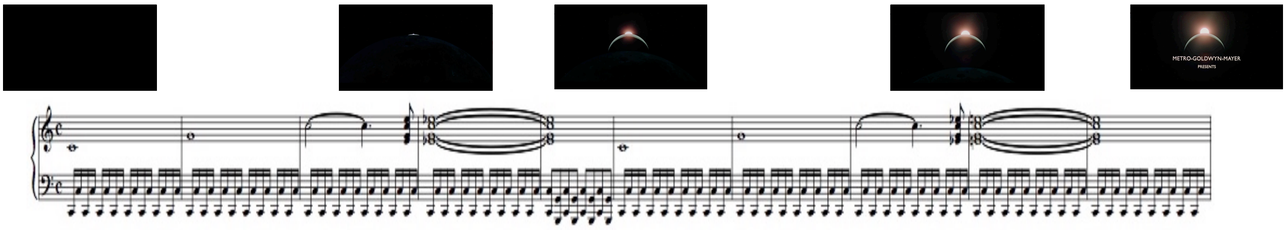
A natural view is that music simply has no semantics, and that it is a formal system that does not bear any relation akin to *reference* to anything extra-musical. A slightly different view is that music has a semantics, but that it pertains to objects that are themselves musical in nature – what we will call an 'internal' semantics. In particular, Granroth-Wilding and Steedman 2014 endow their formal syntax for jazz chord sequences with a semantics that encodes *paths in a tonal pitch space*. In their analysis, surface chords can be assigned syntactic categories that give rise to derivation trees. Each derivational step in the syntax goes hand in hand with a semantic step, which encodes movements in tonal pitch space. Related intuitions are expressed by Lerdahl, who sometimes compares the meaning of music to a journey through tonal pitch space (see Lerdahl 2001).

Importantly, such an 'internal' semantics is not what we wish to argue for. Rather, we will suggest that musical pieces come with a *bona fide* semantics, albeit a highly underspecified one, and an associated notion of 'truth' in certain (real or imagined) situations. Since this is an unusual view, it should first be motivated on the basis of some examples.

2.2 An example of semantic effects in music

We start with the beginning of Strauss's Zarathustra ('Sunrise') [https://www.youtube.com/watch?v=UH3IULiXo24], which is used as the sound track of the beginning of the movie *2001: a Space Odyssey* [https://www.youtube.com/watch?v=e-QFj59PON4&t=0,14s]. In (1), we have superimposed some of the key images of the movie with a 'bare bones' commercial piano reduction (by William Wallace). The correspondence already gives a hint as to the inferences one can draw from the music.

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

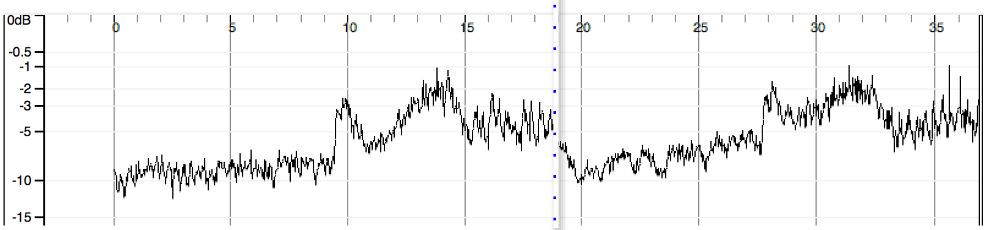


Specifically, the film synchronizes with the music the appearance of a sun behind a planet, in stages – two of which are represented here. Bars 1-5 correspond to the appearance of the first third of the sun, bars 5-8 to the appearance of the second third (4-5 more measures are needed to complete the 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 an antecedent-consequent structure. But the music triggers more subtle inferences as well. A hearer might get the impression that there is gradual development and a marked retreat at the end of the first part, followed by a more assertive development in the second part, reaching its (first) climax in bar 5. Several factors conspire to produce this impression. Three are mentioned in (2). In (2)a, we use chord notation to represent the harmonic development. In (2)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 (2)c we use standard dynamics notation to encode loudness.

(2) a. Harmony: I – V – I – I Maj. chord – I min. chord   I – V – I – I min. chord – I Maj. chord

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

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



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

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

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

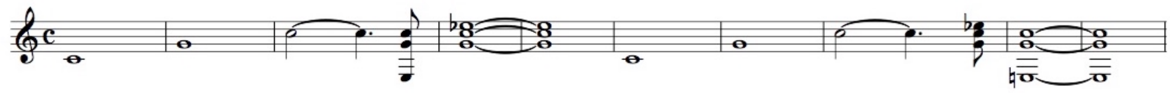
There would definitely be more subtle effects to discuss. But even at this point, it is worth asking whether harmonic and melodic movement are *both* crucial to the observed semantic effect, in particular that the development retreats at the end of the antecedent. The question can be addressed by determining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

One way to test (i) is to remove notes responsible for the upward or downward melodic movement while keeping the harmony constant. This is done on the basis of the very simple piano reduction in (1), further simplified to (3)a. In (3)b, two (highlighted) E's responsible for the melodic movement were removed. The initial effect (unstable ending at the end of the antecedent, stable ending at the end of the consequent) is still largely preserved. This might in part be due to the fact that the harmonics of the remaining E's produce the illusion of the same melodic movement as before. But the semantic effect observed is arguably weakened when these remaining E's are lowered by one octave, as is seen in (3)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 plays a role as well.

(3) a. A 'bare bones' piano reduction of Strauss's Zarathustra, measures 5-13 (= same as the reduction in (1), without lower voice)
https://soundcloud.com/philippeschlenker/strauss-zarathustra-2-10-standard-no-base?in=philippeschlenker/sets/prolegomena-to-music-semantics

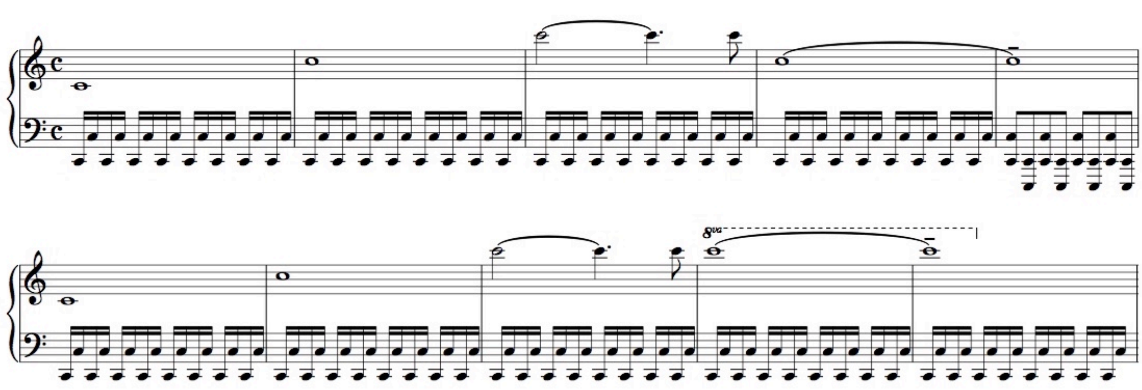
b. Same as a., but removing notes responsible for the downward or upward movement of the soprano in a. (the notes that were removed are appear in red in a.)
https://soundcloud.com/philippeschlenker/strauss-zarathustra-2-10-no-mel-mvt-no-base

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



The contribution of the melodic movement can be further highlighted by turning to (ii) and asking what effect is obtained if we rewrite (3)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 haven't so much constructed a 'minimal pair' (since there are many differences between (4) and the reduction in (1)) 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.)

(4) A version of (3)a re-written using only the note C
https://soundcloud.com/philippeschlenker/strauss-zarathustra-2-10-all-c?in=philippeschlenker/sets/prolegomena-to-music-semantics



In sum, in this case harmonic and non-harmonic properties conspire to yield a powerful effect, and their respective contributions can be isolated by rewriting the piece in various ways. 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 source of noise can usually be inferred to have more energy if it is louder; and given a fixed source, if the frequency increases, so does the number of cycles per time unit, and hence also the level of energy. On the tonal side, normal auditory cognition won't be directly helpful to draw inferences, but it seems that stability properties of tonal pitch space are somehow put in correspondence with stability properties of real world events.

The challenge in the following sections will thus be twofold. First, we should establish more systematically that inferences are indeed drawn on the basis of normal auditory cognition on the one hand, and of properties of movement in tonal pitch space on the other; we will attempt to do so in Sections 3 and 4. Second, we should develop a framework in which both types of inferences can somehow be aggregated; we will sketch one in Section 5.

3 Inferences from Normal Auditory Cognition

Heider and Simmel 1944 [https://www.youtube.com/watch?v=VTNmLt7QX8E] famously showed that abstract animations involving simple geometric shapes such as triangles and circles can be construed (given certain movements) as involving objects, and more specifically animate entities with goals and even personalities. As they write, "the movements of lines and figures are the stimuli; but these movements become anchored in a field of objects and persons and are interpreted as acts.". Despite its distance from normal auditory cognition, we will argue (following Bregman 1994) that music equally triggers inferences about objects that are posited as 'virtual sources' of the music. The analysis will leave open whether the sources should be taken to be animate or not, something that will be determined on the basis of the particular properties of the music under consideration.

For brevity, we leave out of the present discussion cases of musical iconicity, in which a musical sequence resembles auditory properties of an event it evokes (for instance, a clarinet off-stage is used to evoke a cuckoo by way of a series of descending two-note sequences in Saint-Saëns's *Carnival* [https://www.youtube.com/watch?v=5LOFhsksAYw&t=10m35s]; while cannons and the Marseillaise are used to evoke retreating French armies in Tchaikovsky's *1812 Overture* [https://www.youtube.com/watch?v=ZrsYD46W1U0&t=12m39s]). Suffice it to say that these inference are immediately captured by a source-based semantics, and thus that non-iconic effects are of greater theoretical interest.

3.1 Sound and silence

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

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

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

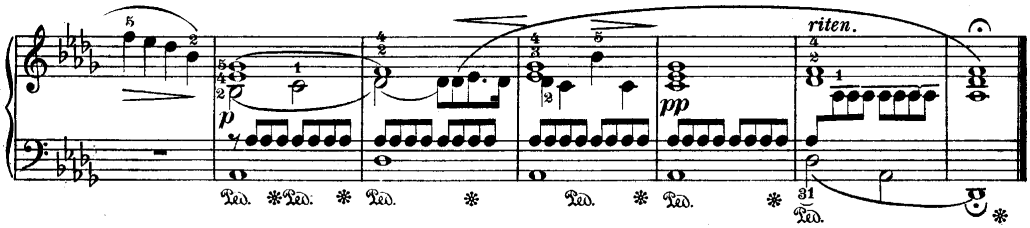
3.2 Speed and speed modifications

Since sound (as opposed to silence) provides information about events undergone by the source, changes in the speed of appearance of sound will be interpreted as changes in the rate of appearance of the relevant events. In the quoted piece on kangaroos (in (5)), each series of jumps starts slow, then accelerates, and then ends slow – and 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 [https://www.youtube.com/watch?v=5LOFhsksAYw&t=3m21s]. It features an extremely slow version of a famous dance made popular in an opera by Offenbach (the 'infernal galop' [https://www.youtube.com/watch?v=5LOFhsksAYw&t=3m21s]). Saint-Saëns's version evokes very slow moving objects that attempt a famous dance at their own, non-standard, pace. Similarly, Mahler's Frère Jacques [https://soundcloud.com/philippeschlenker/mahler-frere-jacques-1-6-normal/s-cyGju] departs from the 'standard' Frère Jacques not just in being in minor key (and in some melodic respects), but also in being very slow – which is important to evoke a funeral procession. A version of a midi file in which the speed has been multiplied by 2.5 [https://soundcloud.com/philippeschlenker/mahler-frere-jacques-1-6-speed25/s-OG4VF] loses much of the solemnity of Mahler's version (and it also sounds significantly happier – see Schlenker 2016 for discussion).

There are certainly more abstract effects associated with speed. In our experience of the non-musical world, speed acceleration is associated with increases in energy, and conversely deceleration is associated with energy loss. This is probably the reason why it is customary to signal the end of certain pieces with a deceleration or 'final ritard' (see Desain and Honing 1996). An example among many 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 doesn't entirely remove the impression but just weakens it).

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



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

In addition, sources that are analyzed as being animate can be thought to observe an 'urgency code' by which greater threats are associated with faster production rates of alarm calls (e.g. Lemasson et al. 2010). This presumably accounts for the association of greater speeds with greater arousal, although this would require a separate musical and ethological discussion.

3.3 Loudness

A noise that is becoming louder could typically be interpreted in one of two ways: either the source is producing the noise with greater energy, or the source is approaching the perspectival center (see Eitan and Granot 2006, and Bregman 1994). The first case is pervasive in music. The second case can be illustrated by manipulating the loudness of a well-known example. The beginning of Mahler's (minor version of) Frère Jacques (First Symphony, 3rd movement) starts with the timpani giving the beat, and then the contrabass playing the melody, all pianissimo, as shown in (7)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.

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

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

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

Without any manipulation, the end of Mahler's Frère Jacques displays a decrescendo that can probably be interpreted as the source gradually losing energy, but which can also plausibly be construed as a procession moving away from the perspectival center [https://soundcloud.com/philippeschlenker/mahler-frere-jacques-last-6-normal].

Interestingly, just looking at the interaction between speed and loudness, we can begin to predict how an ending will be interpreted. As noted, a diminuendo ending can be interpreted as involving a source moving away, or as a source losing energy. In the former case, one would not expect the perceived speed of events to be significantly affected. 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 (7); without the ritenuto, the source is easily perceived as moving away.[1]

(8) Last bars of Chopin's Prelude 15 ('Raindrop')
a. Exaggerated version of the diminuendo in the normal version, with a ritenuto
https://soundcloud.com/philippeschlenker/chopin-prelude-15-last-2-bars-normal-dim
The source seems to gradually lose energy, becoming slower and softer.
b. Same as a., but without ritenuto
https://soundcloud.com/philippeschlenker/chopin-prelude-15-last-2-bars-no-rit-dim
The source seems to be moving away, as it gradually becomes softer, without change of speed.

3.4 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 for animals – 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). 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). The relevant inference is put to comical effect in Saint-Saëns's Carnival, where the melody of a dance is played with a double bass to figure an elephant [https://www.youtube.com/watch?v=5LOFhsksAYw&t=5m24s]. The specific effect of pitch, keeping everything else constant, can be seen by comparing Saint Saëns's version (in a midi rendition, as in (9)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 if raised by 3 octaves, we even get, if anything, the evocation of a small source (as in (9)c).

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

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

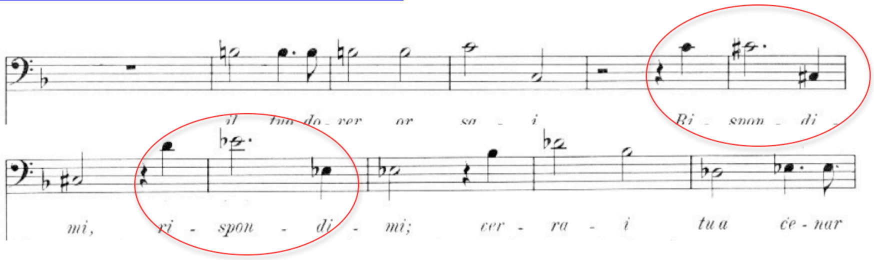
---

[1] If we add a crude crescendo instead, and a final accent, the ending sounds more intentional, as if the source gradually gains stamina as it approaches its goal, and signals its success with a triumphant spike of energy [https://soundcloud.com/philippeschlenker/chopin-prelude-15-last-2-bars-cresaccent]. An intentional, triumphant effect is often produced by fortissimo endings, e.g. at the end of Beethoven's Symphony 8 [https://www.youtube.com/watch?v=C2Avpt9FKP0&t=26m10s].

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

(ii) For a given source, higher pitch is associated with a source that produces more events per time units, hence might have more energy or be more excited. We already saw an instance of this effect in the version rewritten only with C notes of the beginning of Strauss's *Zarathustra* in (1). 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).

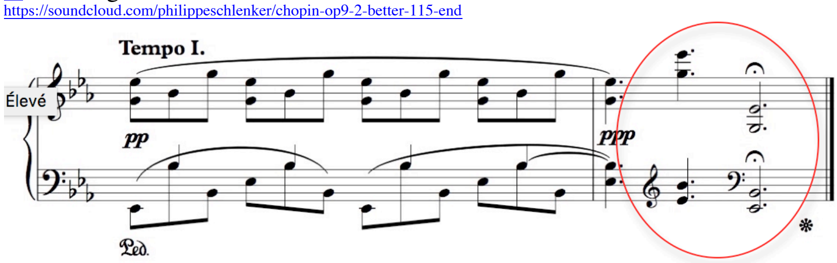
(10) Mozart's Don Giovanni, Commendatore scene, 'Rispondimi': repetition is produced with a chromatic ascent, which contributes to the impression that the Commendatore's request is becoming more pressing
https://www.youtube.com/watch?v=dK1_vm0FMAU&t=3m19s



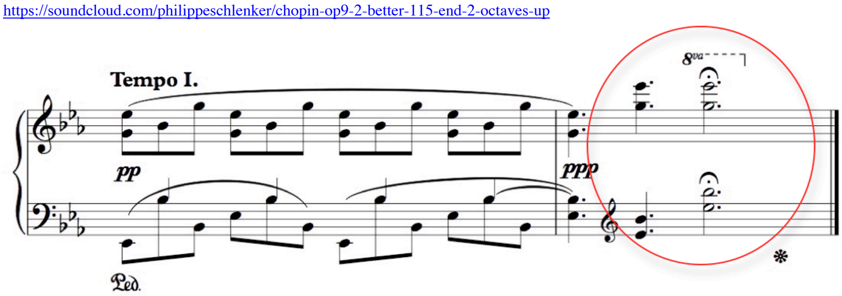
If these remarks are on the right track, all other things being equal, the end of a piece should sound slightly more conclusive if the last movement is downward rather than upward. This effect can be found at the end of Chopin's Nocturne Op. 9/2, which ends with two identical chords, except that the second is 2 octaves below the first. If the score is re-written so that the piece ends upwards rather than downwards, the effect is a bit less conclusive, as is illustrated in (11).

(11) 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.
https://soundcloud.com/philippeschlenker/chopin-op9-2-better-115-end



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



3.5 *Interaction of properties*

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

• The last notes of Mahler's Frère Jacques involve a repetition with attenuation of the loudness, and in a standard version [https://soundcloud.com/user-985799021-177497631/mahler-frere-jacques-last-6-normal/s-qCUPT] they could be interpreted in terms of a source moving away, or gradually dying out. But if a strong rallentando is added [https://soundcloud.com/user-985799021-177497631/mahler-frere-jacques-last-6-ralent/s-qdHOL], the 'moving away' interpretation becomes less likely, and the 'dying out' interpretation becomes more salient – which is exactly the effect we discussed in connection with the end of Chopin's Raindrop Prelude in (11).

• We can also manipulate the beginning of Mahler's Frère Jacques to modify the interpretation of the initial repetitions. A repetition which is realized far more softly that its antecedent may sound like an echo of it, as in (12)b. Louder realization of the repetition may be interpreted as re-assertion, or possibly as a dialogue between two voices, as in (12)c.

(12) Mahler's Frère Jacques (First Symphony, 3rd movement)
a. Beginning, normal version
https://soundcloud.com/user-985799021-177497631/mahler-frere-jacques-3-6-normal/s-10XsM

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

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

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

3.6 Methods to test inferences from normal auditory cognition

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

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

2. Second, minimal pairs should be constructed to assess the inference in a musical context. This could be done in two ways.
(i) One may select actual musical examples, and manipulate them so as to get contrasting pairs, as we did with the end of Chopin's Nocturne 9/2 (in (11)).
(ii) 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 (4)).

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? Or one could resort to indirect ways of testing the inference, for instance by having subjects match musical stimuli with non-musical scenes (e.g. visual ones). Which types of statements will prove most productive is entirely open as things stand, and it is likely that different methods will have to be developed depending on the particular goals of the research. Finally, semantic intuitions might be sharpened by initially restricting the set of models the subjects consider. This 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 – for instance with noise, or in some cases with human voices or even with animal calls – that make it possible to test the parameter under study. In some cases one may even go further and suggest that the relevant properties exist across modalities, and have a counterpart in visual cognition.

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

4 Inferences from Tonal Properties

Lerdahl 2001 makes reference to Heider rand Simmel's (1944) experiment, cited above, "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. Lerdahl argue that similar effects arise in music: "here the dots are events, which behave like interacting agents that move and swerve in time and space, attracting and repelling, tensing and coming to rest". Importantly, for him these inferences arise at least in part on the basis of the behavior of voices in tonal pitch space. Since tonal properties do not have a complete equivalent in normal auditory cognition – unlike loudness, say – we must complement our initial list of inferences with ones that are specifically drawn on the basis of tonal properties. The challenge (to be addressed in Section 5) will be to develop a method to aggregate these heterogeneous inferences.

4.1 An example: a dissonance

A very simple example will help illustrate the inferential power of tonal inferences. In Saint Saëns's impossibly slow version of the *Can Can* dance, which he uses to represent tortoises, there are moments of severe dissonance, and they produce a powerful effect. The very slow dance evokes the tortoises' slow walk. But when we hear a dissonance in measure 12, copied in (13), 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."[2] 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 (13). When the G# is replaced with A throughout this half-measure (as in (13)b), the dissonance disappears, as does the impression that the tortoises are tripping.

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

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

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

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

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

4.2 Cadences

As is well-known, 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 the role of cadences is 'hard-wired' in 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 needs to be stated within a semantic framework. In brief, a full cadence is final because it ends in a position of tonic space that is maximally stable. A half-cadence is less final because it ends in a position that is relatively stable, but less so than a tonic. Furthermore, cadences are often of the form subdominant - dominant - tonic because this provides a gradual path towards tonal repose, which mirrors one of the patterns we saw with speed and loudness, both of which could be decreased gradually to signal the end of a piece. A semantic analysis could in principle capture these facts as follows: music is special (compared to noise) in that the sources are understood to exist in a space with very special properties, isomorphic to those of tonal pitch space. In particular, different positions in tonal pitch space come with different degrees of stability, and 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 has a tonic in root position than if it appears in inverted form. This is presumably because in the former case the chord is more stable.

• If the final I chord is replaced with a VI chord (which shares with it 2 out of three notes – e.g. C E G vs. A C E), the result is less stable – hence the term of a 'deceptive cadence'.

It is worth giving an example of the effect of the slightly 'incomplete' feeling produced by a deceptive cadence. (14)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 (14)b, only the last two bars are changed, and the melodic line is kept constant, but the harmony is modified so as to obtain a sequence V – VI – hence a 'deceptive' cadence. The effect is considerably less conclusive. By contrast, the same kind of modifications have been made in (14)c, except that now the piece ends in a 'plagal' cadence (here: II – I). The effect seems rather conclusive, possibly as much so as the perfect cadence V – I, and certainly much more so than the deceptive cadence V – VI.

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

a. Perfect cadence: II V I

https://soundcloud.com/philippeschlenker/mozart-ah-vous-dirai-je-maman-base

b. Deceptive cadence: II V VI

https://soundcloud.com/philippeschlenker/mozart-ah-vous-dirai-je-maman-va2a-deceptive

c. Plagal cadence: II V IV

https://soundcloud.com/philippeschlenker/mozart-ah-vous-dirai-je-maman-vb2-plagal

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

4.3 Modulations

As is also well-known, tonal pitch space is organized into regions, which correspond to keys – with relations of distance among those. 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 sources: a jump to a distant key would be understood as being odd because it would violate this principle.

A simple example of a spatial interpretation of a modulation can be found in Saint-Saëns's Swan. The title as well as the initial undulating harp accompaniment are evocative of a movement on water – given the title, that of a swan. The piece is initially G Major but modulates to B minor in measures 7-10, as seen in (15)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 (15)b-c.

(15) Saint-Saëns, The Swan, initial modulation (b. and c. re-written by A. Bonetto)
a. Original version, in G major, with a modulation in B minor in measures 7-10
https://soundcloud.com/philippeschlenker/saint-saens-cygne-normal

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

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

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

4.4 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 study them. In the study of inferences from normal auditory cognition (in Section 3), we could (i) select a semantic effect triggered by a certain property X of the music, and (ii) argue that X gives rise to similar inferences with non-musical stimuli. But because of what tonality is, part (ii) is not applicable in the present case. So the analysis must *per force* be more theory-internal. We thus propose that it should include the following steps.
1. First, a hypothesis should be stated – for instance that the leading tone is 'attracted' to the tonic and thus creates an expectation that a voice in the leading tone will then reach the tonic.
2. Second, minimal pairs should be constructed to establish the point. The general methods developed in experimental studies of music could presumably play a role here. In particular, intuitions could be made sharper by restricting the set of models of the music by specifying – by way of a title or a description – what the music is supposed to be about, and then testing semantic inferences that arise given this assumption (this is precisely what Saint-Saëns's title *The Swan* does in the case we just discussed).
3. Third, instead of correlating these effects with ones that are found in non-musical stimuli, one could seek to explain these effects by properties of tonal pitch space as analyzed (on non-semantic grounds) by the best experimental and formal studies.
As mentioned in connection to cadences, semantic analyses will become particularly interesting when

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

5 Musical Truth

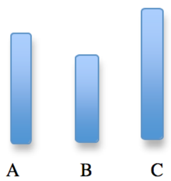
5.1 Inferences and interpretations

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. 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.

Under what conditions will a musical piece be true of a situation? We will take musical events to be auditory 'traces' of events undergone by virtual sources. And we will take a series of musical events to be true of a series of real world events if certain relations among notes or chords correspond to designated relations among events; for instance, a louder note should correspond to a real world event which has greater energy or is closer; a more consonant chord should correspond to a more stable real world event, etc. The basic mechanism can be illustrated in a different domain by considering simplified pictorial representations, seen as visual traces of certain objects. An example is given in (16), where three columns of various heights (A, B, C), arranged from left to right, are used to depict individuals as in the real world scenes in (17), involving a boy, a nurse and business woman.

(16) A pictorial representation



(17) Three possible denotations for (16)

a.

b.

c.

We focus on two relations among the columns that appear in (16): '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 → business woman in the scene (17)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 business woman is taller than the boy, who is taller than the nurse. Thus we can say that on this assignment of values to the columns, the pictorial representation in (16) is true of (17)a. By contrast, it is immediate that the assignment A → nurse, B → boy and C → business woman would fail to preserve the relation 'is to the left of', since (from our perspective) A is to the left of B in (16), but the nurse is not to the left of the boy in (17)a.

• On the assignment A → nurse, B → boy and C → business woman, the relation 'is to the left of' in (16)(17) is preserved in scene (17)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 (16) is not true of scene (17)b (in fact, no assignment of denotations could preserve both 'is to the left of' and is 'taller than' in this case).

• The same type of problem arises for (17)c on the assignment A → business woman, B → boy and C → nurse: the relation 'is to the left of' is preserved, but the relation 'is taller than' isn't, because C is taller than B but the nurse isn't taller than the boy. Hence on this assignment (16) is not true of (17)c (nor could any assignment preserve both 'is to the left of' and 'is taller than').

We will apply the same type of definition of truth to musical pieces, but with relations that are more abstract than the ones 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'). In the 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 could 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.

5.2 An example of musical truth

Because this is all rather abstract, we should start with a highly simplified example. Think again of the C – G – C progression we saw in Strauss's *Zarathustra*, where it was used to evoke a sunrise. We discussed at some length the role played by pitch height, but here we will focus on just two properties, one harmonic and one not:
• within this initial sequence, the key is C (major or minor – this is initially underspecified), and thus C is more stable than G; as a result, the progression is from the most stable position, to a less stable position, back to the most stable position;
• in addition, the progression is realized crescendo.

In order to analyze progressions that just involve these two parameters, we will consider sequences of pairs of the form <note/chord, loudness>, as illustrated in (18). 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).

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

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

Now we will take each pair of the form <note/chord, loudness> to denote an event in the world. Our musical pieces M, M' and M" will thus each depict a series of 3 events in the world. 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 *objects in the world*. Accordingly, we associate:
(i) with any voice M an object O;
(ii) with the series of musical events $m_1, \ldots, m_n$ that make up M, a series of world events $e_1, \ldots, e_n$, with the requirement that each of these events should have O as a participant.

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

(Wolff 2015 also argue for an event-based analysis of musical meaning. But he argues that the ontology of music involves 'pure' events, which we take to be events without individual participants. Since objects are crucial to many inferences triggered by music, we do not share this restriction to 'pure' events.)

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

Starting from the pieces in (18) and the specification of possible denotations in (19), we will say that the music piece $M = <M_1, \ldots, M_n>$ is true of the pair of an object and events it undergoes, $<O, <e_1, \ldots, e_n>>$, just in case $<O, <e_1, \ldots, e_n>>$ is a possible denotation for M, and in addition the mapping from $<M_1, \ldots, M_n>$ to $<e_1, \ldots, e_n>$ preserves certain requirements, listed in (20).

(20) 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 doesn't need justification, the Loudness and Harmonic stability conditions do.
• The preservation condition on Loudness is disjunctive. The intuition is that in auditory cognition in general, louder sounds are associated either with objects that have more energy, or with objects that are closer, as discussed in Section 3.3.
• 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 4.1 in connection with Saint-Saëns's Tortoises, where a dissonance was rather clearly interpreted as the tortoises tripping.

We can now illustrate how these preservation conditions will lead to a notion of truth. We consider three objects: the sun, a boat, a car. And we will consider 'bare bones' versions of two sequences of events for each. For the sun, a sunrise and a sunset. For the boat, a movement towards the perspectival center, and a movement away from it. For the car, just a car crash. We will analyze these events in a highly simplified fashion, with each event made of three sub-events. In this way, we will obtain five possible denotations for our piece M = <<I, 70db>, <V, 75db>, <I, 80db>> in (18)a.

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

Since M is comprised of three musical events, and each of the sequences in (21) is of the form <object, <event_1, event_2, event_3>>, each is a possible denotation for M according to (19). 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 objet approaching the perspectival center, as in boat-approaching.

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

Let us first note that all the sequences of events given in (21) are intended to obey the time ordering condition stated in (20)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 (21), all we need to check is that it satisfies the Loudness and the Harmonic Stability conditions.

• Consider first sun-rise in (21)a. Since M is crescendo, M_1 is less loud than M_2, which is less loud than M_3. The Loudness condition in (20)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, though in psychological terms the 'energy'-based interpretation seems more relevant). This shows that the Loudness condition is satisfied. Turning to the Harmonic Stability condition, it too would seem to be satisfied: the initial and final 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 (21)b. The Harmonic Stability condition is not the issue: just as with sun-rise, the events that begin and end the process can be taken to be the most static and thus stable. On the other hand, the Loudness condition is not satisfied: when we consider the first and the second event, namely maximal-luminosity and diminishing-luminosity, there is neither an increase in 'energy' level, nor an approach.

• Consider now boat-approaching in (21)c. As was the case for sun-set, both the Loudness condition and the Harmonic stability condition will now be satisfied. While it wouldn't make much sense to say that a boat approaching is gaining energy (if anything, it might slow down as it approaches the coast), the sequence corresponding to boat-approaching = <boat, <maximal-distance, approach, minimal-distance>> satisfies the Loudness condition for M because the events in the sequence are getting gradually closer to the perspectival center. The ordering $M_1 < M_2 < M_3$ of the crescendo sequence M is thus preserved when M is mapped to the series of events boat-approaching, since the sub-events maximal-distance, approach, minimal-distance are similarly ordered in terms of proximity to the perspectival center. Turning to the Harmonic stability condition, the situation is the same as in sun-rise (or for that matter sun-set): the initial and final sub-events are the most static, whereas the intermediate event is more dynamic, and less dynamically stable. This corresponds to the ordering by stability of the I V I sequence in M.

• The boat-departing event in (21)d satisfies the Harmonic stability condition, but not the Loudness condition. Its two initial sub-events are maximal-distance followed by departure, and the second doesn't have more energy than the first, nor is it closer than it – hence the crescendo character of M is not properly interpreted.

• Finally, the car-crash event in (21)e might or might not satisfy the Loudness condition, depending on whether we take the sequence <movement_1, movement_2, crash> to correspond to an increase in energy and/or to a movement towards the perceiver. But plausibly the Harmonic stability condition is not satisfied: one would expect that the musical event corresponding to the crash is the least stable of all three events, whereas here it corresponds to the final I of the piece. Things would be different if the piece finished in a highly dissonant chord, but this is not the case here.

In summary, the piece M introduced above is true of sun-rise and boat-approaching but not of the other events considered here. 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 at it should be: musical inferences are highly underspecified, and this property should be preserved by an adequate semantics.

6 Musical Syntax and Event Mereology

Having argued that one can make sense of the referential content of music, we will now suggest that aspects of musical syntax can be reinterpreted in semantic terms. Specifically, we will 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, not 'syntax'). In particular, we will propose that a musical group A is taken to belong to a musical group B if the real world event denoted by A can naturally be taken to be 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'.

6.1 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 in Pesetsky and Katz 2009 for prolongational reductions. By contrast, we will be solely concerned with grouping structure and time-span reductions. Lerdahl and Jackendoff's own discussion departs in two respects from a 'generative syntax' analysis.
(i) First, they take their structures to be based on parsing rather than generation, and to be rely heavily on preference principles rather than categorical principles of well-formedness.
(ii) Second, Lerdahl and Jackendoff take some of their own structures to be based in perception and to follow from very general Gestalt principles.
(i) may or may not be essential, for one might present the same system in terms of parsing or generation, as Pesetsky and Katz 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*.

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

6.2.1 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 (22); it has in particular the consequence that if an event e has parts, then *their* parts are also parts of e (Transitivity).

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

The notion of 'proper part' follows from that of a part: x is a proper part of y iff x is a part of y and x and y are not identical. For simplicity, we will further assume that every event is made of atomic events, i.e. events that do not themselves have proper parts, as defined in (23).

(23) 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

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

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

First, it makes sense to assume that atomic events are ordered in time, as stated in (25).

(25) 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 (26).

(26) Possible decompositions of abc - simplified notation

a. abc –> a, b, c

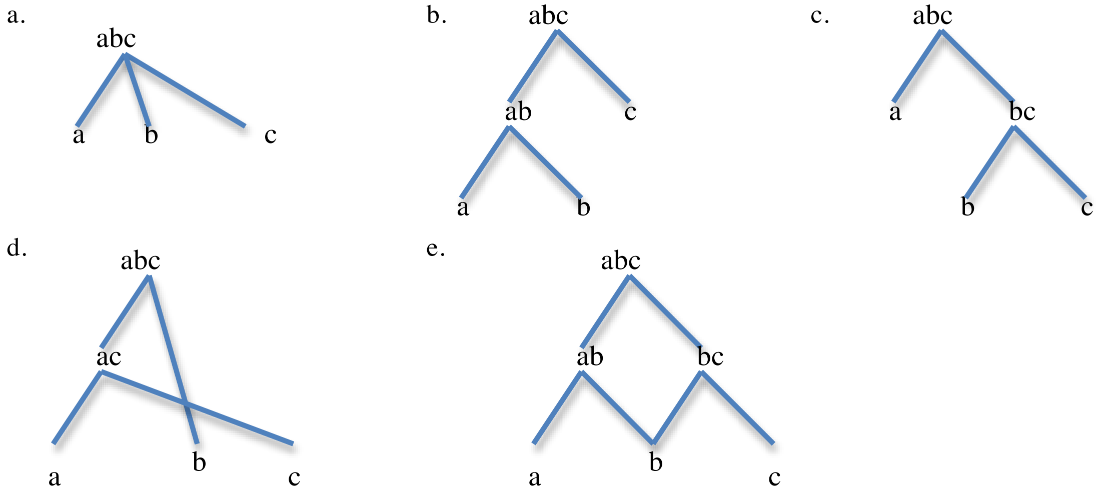
b. abc –> ab, c

c. abc –> a, bc

d. abc –> **ac**, b

e. abc –> **ab, bc**

Now it can immediately be seen that (26)a, b, c correspond to 'standard' 'syntactic' trees that could be obtained from a context-free grammar, as illustrated in (27)a, b, c. But (26)d, e require 'trees' with an unusual shape, as illustrated (27)d, e.

(27)


• The situation in (27)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 $D_1 < \ldots < D_n$, with a requirement that if $D_i < D_k$, then all the terminal nodes dominated by D_i precede all the terminal nodes dominated by D_k (see Kracht 2003 p. 46); precisely this condition fails in (27)d, as we can neither have ac < b nor b < ac.

• The situation in (27)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 (27)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 (28), which prohibits overlap among events unless one is contained within the other.

(28) Minimal part-of structures

A part-of structure is minimal if whenever x is part of y and x is part of z, y is part of z or z is part of y.

This condition is of course violated by (27)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 (real world) events are best decomposed in a non-economical fashion, with a part which is common to both (for instance because there is a smooth transition between two events[3]);
(ii) occlusion: cases in which there is a reason to think that two distinct events share the same auditory trace.
We come back in Section 6.2.2 to exceptions of both sorts.

• Consider now (27)d. It leads one to posit that an event has a discontinuous auditory trace. Two assumptions are needed to prohibit this case.
(i) The first assumption, which makes much intuitive sense, is that real-world events are normally connected. But this measure is not enough. Consider an analogous case in the visual domain. It makes sense to posit that both objects and events satisfy a condition of spatial or temporal connectedness. Still, due to occlusion, there are numerous objects and events that we *see* as disconnected, even when our cognitive system is able to take occlusion into account and to posit a single underlying object or event despite the disconnected nature of the percept.
(ii) Thus in order to prohibit structures such as (27)d we must also posit that cases of auditory occlusion don't occur. This makes much sense in some standard situations: if you are in the middle of a conversation while a car passes by, it will rarely happen than the background noise is so strong as to occlude the conversation, or conversely.

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

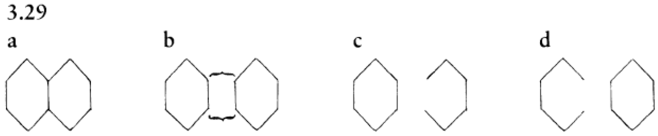
6.2.2 Exceptions

Lerdahl and Jackendoff 1983 emphasize that cases such as (27)e 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. Furthermore, the exceptions they list are of the two types we announced above: in case so overlap, the denoted events are construed as sharing a part; in cases of occlusion, the auditory trace of an event occludes that of another event.

❑ *Overlap*

Lerdahl and Jackendoff 1983 illustrate visual overlap by the case in which a single line serves as the boundary between two objects, and is thus best seen as belonging to both, as in (29)a, which is preferably analyzed as (29)b rather than as (29)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.

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



Lerdahl and Jackendoff 1983 cite the very beginning of Mozart's K. 279 sonata as an example of auditory overlap, as seen in (30). The I chord at the beginning of bar 3 seems to both conclude the first group and initiate the second, hence it can be taken as the trace of an event that plays a dual role as the end of one event and at the beginning of another. Alternatively, and less plausibly perhaps, this 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 (29)).

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

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

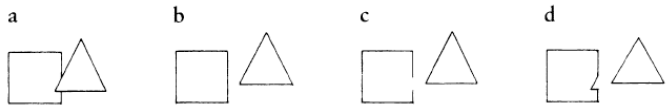
3.25

❑ *Occlusion*

The second case involves part of an object occluding another object, as in (31). Here the most natural interpretation of (31)a is as (31)b, which involves occlusion, rather than as (31)c-d, which don't.

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

3.30



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

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

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

3.26

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

6.2.3 Sequencing events

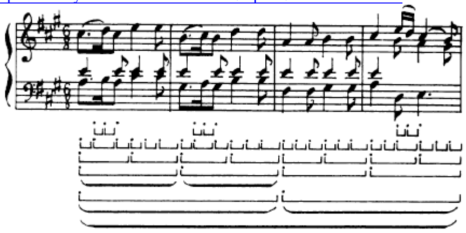
For our analysis to be plausible, one would need to establish that *independently from music* (or language, for that matter), events are naturally perceived with a part-of structure. Jackendoff 2009 argues that there are tree-like structures outside of language, and he gives the example of actions, which 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. 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.

6.3 Time-span reductions and headed events

It is uncontroversial that Western classical music has a metrical structure that yields an alternation of strong and weak beats. Lerdahl and Jackendoff 1983 analyze it with rules that are very similar to those used in metrical phonology. They take metrical structure to be essential to the organization of events at micro-levels. At larger levels, they take them to be organized by grouping structure. But Lerdahl and Jackendoff argue the structures obtained are still insufficient in that they fail to distinguish different levels of importance within musical groups. Formally, they propose that their tree structures should be seen *headed*: in each natural unit, one musical event 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. In Lerdahl and Jackendoff's words,

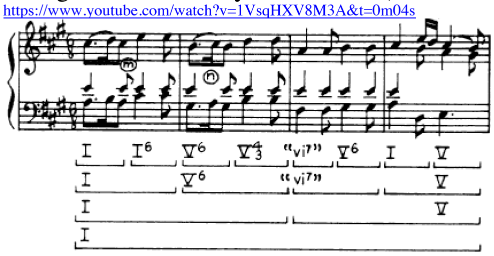
> at the most local levels, the metrical component marks off the music into beats of equal time-spans; at larger levels, the grouping component divides the piece into motives, subphrase groups, phrases, periods, theme groups, and sections. Thus it becomes possible to convert a combined metrical and grouping analysis into a time-span segmentation, as diagrammed for the beginning of [Mozart's] K. 331 in 5.11. (p. 119)

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



The next step in the construction of time-span segmentation is the selection of a head in each group, as is illustrated in (34).

(34) Time-span reduction obtained from (33) by selecting in each the musical event which is metrically strongest/harmonically most stable (Lerdahl and Jackendoff 1983)
https://www.youtube.com/watch?v=1VsqHXV8M3A&t=0m04s



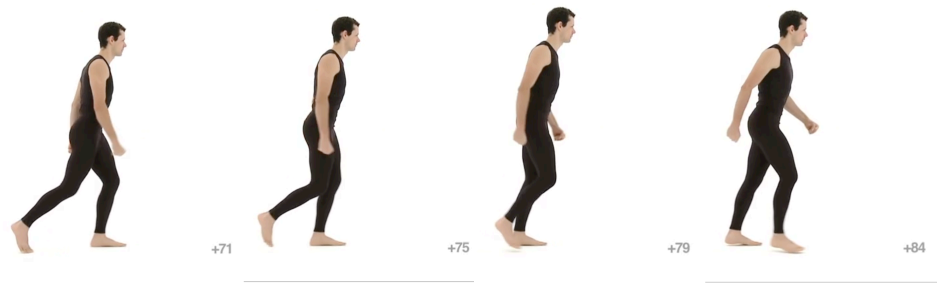
As Lerdahl and Jackendoff write (p. 120), "in the span covering measure 2, the V^6 is chosen over the $V^4{}_3$, and proceeds for consideration in the span covering measures 1-2.; here it is less stable than the opening I, so it does not proceed to the next larger span; and so forth. As a result of this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

It remains to ask whether the headed nature of time-spans should be taken as primitive, or 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 (35). 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 'head' of the relevant event, in terms of the present discussion. These are clearly points at which impulses of energy are given, somehow like points of metrical strength in music, and probably also points of greatest physically stability.

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



7 Conclusion and further questions

7.1 Theoretical conclusions

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

We further argued that aspects of musical syntax can arguably be reconstructed on semantic grounds. In particular, 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.

7.2 Methodological conclusions

Although we based our theoretical discussion on informal introspective judgments (which could be subjected to experimental methods in the future), we made frequent use of 'minimal pairs' to display semantic effects. Once a potential semantic effect was identified, and a hypothesis formulated as to its origin, the analysis could be tested by isolating the crucial parameter, in one of two ways: in most cases, we discussed minimal musical pairs that differed whenever possible by just one parameter, the crucial one, in order to show that the target semantic effect was weakened when the parameter was distorted; in a couple of cases, we were able to abstract away (rather than control for) other parameters, as when we 'rewrote' a passage with C's only in order to display the specific effect of (non-harmonic) melodic movement.

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

for instance by providing additional information (by way of titles, stories, or other non-musical information) so as to make inferences more precise.

7.3 Further questions

Several important questions are left for future research.

(i) Since we only attempted to provide a 'proof of concept', we sketched the barest of music semantics. Real applications would of course require a more sophisticated analysis, but also a determination of the fine-grainedness of the semantics under study. As a point of comparison, on a very coarse-grained picture semantics, a hexagon may count as a map of France, but this would not be the case on a more-grained semantics; similar issues of fine-grainedness will arise in music semantics.

(ii) The meaning of music is usually construed in terms of *emotions*. These do not play a primitive role in the present account. Still, the objects corresponding to the pseudo-sources of the music may be perceived as emotional, just as triangles and circles in Heider and Simmel's (1944) experiment. Thus the fact that emotions are not primitives of the present account does not imply that they do not have an important role to play in music semantics. (In Schlenker 2016, we make the further suggestion that the formal framework could be minimally modified to replace 'real world events' with 'experienced events', in such a way that tonal notions of tension and relaxation could yield a more direct mapping to emotions.)

(iii) We have been silent about music *pragmatics*, the set of non-semantic inferences drawn by reasoning on the motives of the intentional subject that produces a message, be it linguistic or musical. In Schlenker 2016, we argue that some pragmatic effects, pertaining to information packaging, may arise in the absence of a semantics, while others might interact in interesting ways with the present account.

(iii) As things stand, our account has nothing to say about aspects of musical syntax that are not captured by grouping structure and time-span reductions. In particular, we leave for future research a potential semantic study of Lerdahl and Jackendoff's (1983) 'prolongational reductions', which play a central role in Pesetsky and Katz's analysis (2009) of music syntax.

(iv) 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.

References

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

Schlenker, Philippe. 2016. Prolegomena to Music Semantics. Manuscript, Institut Jean-Nicod and New York University.

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

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

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

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

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