Walk–denoting music: refining music semantics

Léo Migotti[1] and Léo Zaradzki[2]

[1] Institut Jean Nicod, Département d'études cognitives, ENS, EHESS, CNRS, PSL Research University, Paris, France

[2] Laboratoire de Linguistique Formelle, Université Paris Diderot, Paris, France

Abstract

Music has recently been argued to have a referential semantics (Schlenker (a,b)), *i.e.* to trigger inferences about an extra–musical reality. In this view, because the set of possible denotations is often very large, the meaning of music is often very abstract. Here we consider a very particular kind of musical sequences, which we call walk–denoting as they strongly evoke walking–situations — namely situations in which at least one character is walking. We show that the current model for music semantics is doubly insufficient. First, it makes wrong predictions with respect to the considered musical snippets. Using the method of minimal pairs, we come up with an enhanced model that accounts for inferential differences that the previous one left aside. Second, it relies on the non–trivial assumption that all notes are interpreted as events, while alternative theories seem to be just as plausible. Because a rewriting of our prototypical musical snippet adding a quaver did not seem to affect the denotation, the possibility that some musical events denote nothing needs to be investigated. Finally, we sketch the overall theoretical landscape through two main theories, which either consider that all musical events are interpreted, or that some of them might not be.

1 Introduction

Recent investigations about the application of formal linguistics methods to non–linguistic objects such as music strongly suggest that music can convey information about the world through semantic rules that bridge the characteristics of music and the ones of what it can evoke or represent. While evidence has been provided regarding the interpretation of some musical features from a purely semantic perspective (Schlenker (a,b)), little attention has been given to the systematic link that exists between the internal structure of music, *i.e.* its syntax, and the information it conveys, *i.e.* its semantics. Yet, we know that both music and the situations evoked can be represented hierarchically (Jackendoff (2009); Schlenker (b)). It then seems to make much intuitive sense to posit that if music has a semantics, its syntax, in relation with that of the denoted situation, has to play a role as well. We first present Schlenker's theory of musical semantics. We discuss a case–study about musical snippets evoking walking–situations, and we then highlight some of the limitations of Schlenker's model. We argue that Schlenker's theory lacks conditions on the rules linking music and situations structures. We finally present two possible theoretical accounts of this relationship. We will keep the choice to be made for further theoretical and experimental research.[1]

[1]We would like to deeply thank Philippe Schlenker for providing us with reliable introspective judgments as well as crucial theoretical insights, and for reviewing this article. We also thank Emmanuel Chemla who helped us clarifying our theoretical hypotheses. Thank you also to all our informants for taking the time to describe the detailed spontaneous inferences they drew from our musical stimuli.

2 A formal model for music semantics

Because music can evoke or make us think about certain events, be they real or not, and describe some situations better than others, Schlenker (a,b) argued that it must have a semantics. Indeed, music does not only convey information about its form and its internal structure, it also conveys information about an extra–musical reality: some music evoke sad or happy situations, others might well describe a landscape or an animal. The set of all the hypothetical situations a music can appropriately describe is therefore taken to be its meaning. The following section draws from Schlenker (a,b) and presents the concepts, terminology and notation that are needed for our theoretical proposals in section 4.

The first core idea we rely on is that music is able to convey information about the world because certain musical parameters such as timber, pitch, loudness or harmonic stability (among many others) are semantically interpreted: each of them bears some of the meaning of music (Schlenker (a,b)). For instance, pitch might well provide information regarding the size of a character involved in a scene that is depicted by the music. From now on we will talk about *inferences* to refer to this information music provides and listeners get. Also, we will call what these inferences are about *virtual sources*[2].

Because music shares many features with sound, but cannot be reduced to it, the musical parameters responsible for the triggering of inferences can be split into two main categories. The first category gathers the parameters music and sound have in common, which makes it possible to derive semantic rules from normal auditory cognition. Loudness is of this sort: as any sound in nature has a certain level of loudness, we know from our world experience that loudness can be linked to certain properties of the actual source of the sound, arguably either its distance to the listener, or its level of energy. Applying this to music, we get a semantic rule on loudness interpretation, according to which the loudness of any musical event is ambiguous and either interpreted in terms of distance or energy of the *virtual* source. The second category gathers parameters that have no trivial counterpart in the non–musical world, and are intrinsically linked to tonal and harmonic properties of music. For instance, the harmonic stability of a given chord in a given key follows from tonal rules. As we do not experience harmonic stability of non–musical events, we cannot derive a semantic rule from auditory cognition; rather, Schlenker argues that this parameter is interpreted as the *actual* stability of the source, or that of the emotional state in which the listener is put.

The second core idea is that our music semantics needs to state rules linking the musical parameters to their semantic interpretation. From now on, we will use *'musical event'* to refer to any note or chord, and *'denoted situation'* to refer to any complex situation pertaining to the set of situations a musical snippet can evoke. Formally, we define, just as Schlenker, a musical snippet as an n-tuple $M = (m_1, ..., m_n)$, and a possible situation S as an n-tuple $S = (e_1, ..., e_n)$, where $(m_1, ..., m_n)$ is the succession of notes, each of which represents the corresponding event carrying the same index in the situation.

In formal linguistics, the common view is that the meaning of a sentence is the set of all situations of which the sentence is true. Transposing this to music, the meaning of M is the set of all situations which M is true of. We thus need a notion of musical truth. We say that M *denotes* S ($M \models S$) if S is one — among many others — possible denotations for M. The final step is therefore to find rules to compute the truth–value of a music, given a specific situation it might denote. Schlenker posits that those rules are order–preservation rules: for M to denote

[2]Their naming must not, however, confuse us about their nature: virtual sources are not actual sources of the sound; they are virtual objects that may or may not produce sounds — if they do, the music does not even need to match this sounds — involved in the denotation of music.

S, musical parameters involved in each event in M need to be ordered in the same way as their interpretation in the denotation. For instance, loudness levels and corresponding levels of energy or distances from the listener must be ordered in the same way.

Although the above model makes clear intuitive sense, we argue that it makes wrong predictions regarding the possible denotations of some musical snippets. Specifically, we claim that:

1. It makes incorrect predictions regarding the possible denotations of some specific musical snippets we think trigger strong inferences about a virtual source walking, as shown in next section.
2. It relies on the assumption that each musical event is systematically interpreted, regardless of its structural role, while it seems reasonable to posit that some musical events are more important than others.

3 Walk–denoting music and walking–situations

As stated, the above theory fails to account for some strong inferences we believe to be triggered by the prototypical musical snippet about walking-events in Figure 1[3].

A walking–situation is a situation in which at least one of the virtual sources is walking. A walk–denoting music is a musical snippet that can denote a walking–situation. For levels of stability to match, and a music to denote a walk, we thus need to have a musical event to walking–event matching as shown in Figure 1: bass notes represent footsteps, while second and fourth chords represent the 'bounces' occurring during the transition from one foot to the other.

Based on our own introspective judgments, as well as on that of informants, we argue that the music contained in Figure 1 triggers very strong inferences about walking–situations. One might argue that listeners get these inferences because they are constantly experiencing walking–situations. This argument does not explain, however, the existence of contrastive judgments between musical snippets which, based on Schlenker's model, should all be able to denote a walking–situation, while the inferential judgments we got from our informants do not match the theoretical predictions from the model. Let us consider the score in Figure 2[4].

In order to compute the possible denotations of this piece, we first need to understand what the meaningful parameters are, both in the music itself, and in the virtual walking–situation. Intuitively, we argue that the most prominent parameter which is involved in a walking–situation and varies throughout it is stability: each footstep appears to be a quite stable event, while the transition from one foot to the other are relatively less stable, be it only because a foot is lifted

Figure 1: Prototypical walk–denoting music

[3]audio file: https://www.youtube.com/watch?v=BWbqZ1BiRzI&feature=youtu.be
All scores are directly clickable to access the audio file.

[4]audio file: https://www.youtube.com/watch?v=SvmXWX_xAeU&feature=youtu.be

Figure 2: A music failing to denote a walking–situation

Figure 3: Violating alternation condition

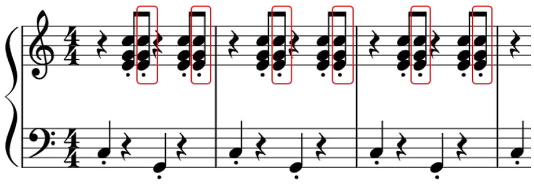

Figure 4: Prototype with extra quavers

in the air. Thus, the corresponding musical parameter *in music* must be harmonic stability. Based on rules of preservation of ordering, the theory predicts that music in Figure 2 must be able to denote a walking–situation, which it does not, according to our introspective judgments. Thus, we need to refine the formal theory that made this incorrect prediction possible.

In order to do so, we varied a whole set of musical parameters once at a time, and selected the most relevant ones according to a few informants. This led us to hypothesise that, for a musical snippet to denote a walking–situation, it has to involve the steady repetition of two different chords, that are both intrinsically stable, and sufficiently close to each other in the tonal space.

A way to test whether this set of conditions is accurate was then to build minimal pairs, *i.e.* couples of stimuli made of the above prototype and a composed musical snippet based on this very prototype but violating one of the five above conditions. Our prediction is thus that violating any of the condition would trigger an inferential preference for the prototypical snippet, that satisfies all conditions[5].

From a theoretical perspective, it appears that these five parameters can be classified in two groups that make cognitive sense, and that can be derived from theoretical considerations. As a walk itself is defined as the alternation of two footsteps, a first natural class of conditions follows from the fact that any music that denotes a walking–situation must also be composed of exactly two events. The second and third conditions on repetition and regularity can be derived from the same physical fact: a normal, stereotypical walk is necessarily the repetition of footsteps (which explains why the musical events must themselves be repeated), and that repetition needs to be approximately symmetric (which explains why the repetition of musical events shall never be broken and remain steady). A second class of conditions appears to be linked to the fact that the right footstep is necessarily different from the left one, but that both events are not so different and are also both rather stable, although one might be a little bit more stable than the other; thus, the corresponding two musical events must be minimally different as well.

These conditions being stated, we were however concerned with the music snippet in Figure 4[6]. Indeed, introspective judgments given by our informants as well as our owns suggested that the denotation of this rewritten version of the prototype, in which a quaver was added on each offbeat, was not affected; or that is was affected in a very subtle way, that did not correspond

[5]In order to check our theoretical intuitions, we are currently running an experiment which aims at checking whether these conditions actually play a role in the triggering of inferences about walking–situations in listeners, by presenting participants the minimal pairs that are available at https://www.youtube.com/watch?v=O4Puddu3wXQ&feature=youtu.be.

[6]audio file: https://www.youtube.com/watch?v=gR-wHm4nFYk&feature=youtu.be

to a radical change in the denotation or a new event. We thus wondered how this could be the case, from a theoretical perspective.

4 Theoretical refinements

In addition to showing that Schlenker's interpretive rules are incorrect or at least underspecified, walk–denoting music examples raise one fundamental question. The example in Figure 4 suggests that some notes may not be interpreted as concrete events in the denoted situation[7]. If this is so, then we have to answer two new questions:

1. If a note is not interpreted as a concrete event, what is its semantic role?
2. How do we determine which notes must be interpreted and which may remain uninterpreted?

In the rest of the article, we discuss theoretical issues related to these questions and present competing theories of musical meaning.

4.1 Hierarchical structures

We know from Lerdahl and Jackendoff (1983) that there are ways to assign hierarchical structures — *e.g.* tree–like structures — to musical pieces. Many different views address the question of which structure is the best to account for musical hierarchy and dependencies, such as time–span reduction or prolongational reduction from Lerdahl and Jackendoff (1983). From a different perspective, Harasim, Rohrmeier and O'Donnell argue that the internal and harmonic structure of any given tonal piece can be accounted for with the three notions of prolongation, preparation and substitution (Harasim et al. (2018)). On the other hand, it has been proposed that events too feature a hierarchical structure that can be represented as a tree (Jackendoff (2009)).

In a famous series of experiments where they asked subjects to segment taped common situations in sub–units, Zacks, Tversy and Iyer showed that people conceive situations as partonomic hierarchies (Zacks et al. (2001)). In particular, people form mental groups of events that recursively embed into one another, often in a goal–directed fashion. On the musical side, Lerdahl and Jackendoff gave formal rules derived from Gestalt principles to determine a so–called *grouping structure* for musical pieces. This grouping structure is exactly of the same partonomic kind as those evidenced by Zacks and colleagues. For example, Figure 5 (taken from Lerdahl and Jackendoff (1983)) shows the grouping structure that the system derives, in accordance with

Figure 5: A possible grouping structure for the opening theme of Mozart's 40th Symphony.

[7]While this is a constructed example, examples exist in music. To give only a famous one, in the opening of *Peter and the Wolf* by S. Prokofiev, after the first occurrence of Peter's theme in first violins, it is played again by the seconds, while the firsts play high–pitched offbeats. We believe that it does not add any event to the denoted situation (*e.g.* Peter gamboling through the meadows) but only gives it a more carefree character. The example can be heard here: https://www.youtube.com/watch?v=uwKgH8QH_mc&t=2m39s.

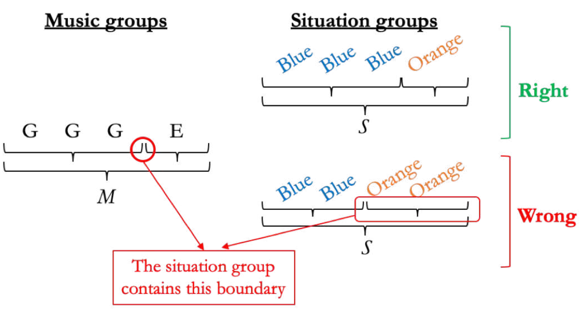

Figure 6: Example of grouping mismatch.

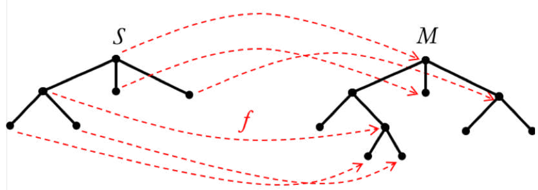

Figure 7: Example of tree embedding

listeners' common intuitions, for the opening theme of Mozart's 40^{th} Symphony. As shown by the curly brackets, the first three notes form a first group together, and the next three form another group. These two groups are then grouped together at a higher level to form a new group, and so on.

Elaborating on ideas already suggested in Schlenker's work, we posit that for a musical snippet M to denote a situation S, the grouping structures of both must match, in the following sense: a group on the situation side must not contain a group boundary on the musical side. More precisely, if $e_1, ..., e_k$ and a are events of S associated in M with musical events $m_1, ..., m_k$ and b, and if there is a group to which $e_1, ..., e_k$ belong but a does not belong, then there must be no musical group to which b together with some but not all the m_i belongs. As an illustration of this phenomenon, let us take the opening motif of Beethoven's 5^{th} Symphony. This consists in four notes: G G G E. According to Lerdahl and Jackendoff's rules[8] they are grouped as [[G, G, G], E]. In the corresponding section of *Fantasia 2000*[9] each note is illustrated by a coloured lightning: the Gs are interpreted as blue lightnings, and the E as an orange one. It seems to us, though, that image and sound fit far less well if we colour, say, the third lightning in orange too. As shown in Figure 6, this is because the orange group would then contain the G/E boundary.

While we believe that preserving the hierarchical structures is a rather natural requirement for the interpretive rules, we will not argue further here, and leave this study for future research. We here assume the following strong version of preservation. First, we posit that musical snippets are associated with a hierarchical structure which is mathematically implemented as a directed rooted tree, possibly with vertices labeled as heads at each level[10]. Such a modeling is compatible with many approaches to musical syntax, either based on grouping structures or rather on harmony. Thus, we do not need to commit to any particular formal system here. Second, in line with Jackendoff (2009), we posit that situations too are associated with a hierarchical structure implemented as a directed rooted tree with heads. While this is a stronger assumption than what had been experimentally proved by Zacks *et al.*, we take it as a working hypothesis. Third, we claim that a necessary condition for a musical snippet M to denote a situation S is that the tree of S can be embedded into the tree of M. Formally, if M is a tree (V, E) and S is a tree (V', E'), where V and V' represent the sets of vertices and E and E' represent the sets of edges[11], a necessary condition for $M \models S$ is the existence of an injective

[8] and also to general Gestalt principles

[9] an animated film where image is intended to be a denotation of the music; see: `https://www.youtube.com/watch?v=nMnlxYkZKaU`

[10] We discuss this notion of heads in more details below. At first glance, let us say that heads are musical events that are more prominent or structurally more important than non–heads.

[11] Since the tree is directed, edges are ordered pairs of vertices.

root–preserving function $f : V' \to V$ such that: $\forall x, y \in V',\ (x, y) \in E' \Rightarrow f(x) \sim f(y)$, where $f(x) \sim f(y)$ means that there exists a path in M from $f(x)$ to $f(y)$. What f does is that it takes any event of S and maps it onto a musical event meant to be its musical representation, in an injective way. Moreover, if an event of S is subordinated to another, then the same subordination relationship should hold between their musical representations. Figure 7 shows an example of such a function.

4.2 Do all notes have the same semantic status?

We now turn to the question of uninterpreted notes. The discussion from the end of section 3 suggested that all notes of a musical snippet are not necessarily interpreted as events in the denoted situation. While we do not have data to decide whether or not this can be, we will here present two competing positions about it.

In Figure 4 we saw that in the walk–denoting case, bass notes were interpreted by steps and other beats by bounces; as for the additional offbeats, it was not clear. What we can say is that most important musical events are bass notes, the other beats are musical 'bounces' of these bass notes, and offbeats are kind of squared bounces (bounces' bounces). Thus, most important musical events are interpreted as important events in the situation too. This suggests that the musical salience of notes plays a role. This is why we added heads to our tree–implementations: heads are special vertices meant to represent the most salient events of a musical snippet or a situation. We will examine below alternative ways of implementing salience and how it can be computed. Now we sketch two informal opposite theories, to be refined below:

- **Theory A** (strong[12] theory): Every note is necessarily interpreted in the situation — though maybe by a very abstract event — that is, every note matches to an event in the situation (in this regard, this is the closest position to Schlenker's). Moreover, more important musical events should correspond to more important events in the situation[13].
- **Theory B** (weak[12] theory): Each note is not necessarily interpreted as an event in the situation. This does not mean that these uninterpreted musical events are semantically vacuous, but they do not refer to a concrete event, rather they modify the denoted situation as some adverb would in linguistics. Yet the musical heads always need to be interpreted — though not necessarily by head events. Moreover, more important interpreted musical events should correspond to more important events in the situation[13].

As an illustration, Figure 8 shows possible trees associated with the music of Figure 4 and with the corresponding walking–situation, and an embedding of S into M as described by Theory B (for the sake of readability, we drew the arrows only for the first branch of S, but the same exist for the other branches). As one can see, not every note in M is matched to an event in S, but every head note is.

As mentioned above, both theories, though opposite, deal with the salience of notes or events. Salience can be implemented in our treeish framework by the mean of heads. These heads may be obtained by formal rules, exactly as the hierarchical structure is. However, we think that this binary notion of heads is too coarse. We may wish, for instance, to distinguish between three degrees of salience or more. We may then add, for instance, an *ad hoc* notion of secondary heads, and require them to be interpreted, but not by first–order heads, and so

[12]The names *weak* and *strong* theories reflect the fact that Theory A requires more constraints.

[13]This last idea is related to what Schlenker has suggested in Section 8.5 of Schlenker (b). There are stronger or weaker ways to express this condition; we will be more precise below. Note that it is also possible to completely drop it.

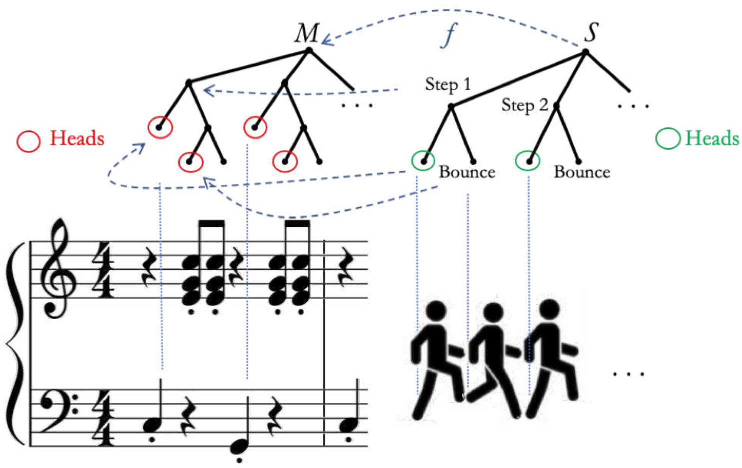

Figure 8: An illustration of Theory B on the example from Figure 4.

on. Since the number seems to be unbounded, we rather replace the notion of heads by that of *weight*[14]. We then suppose that M and S are rooted directed tree coming together with weight functions. A weight function on M (resp. S) is just a function $p : V \to \mathbb{R}_+^*$ (resp. $p' : V' \to \mathbb{R}_+^*$). How these are obtained may rest on formal rules akin to those developed in Lerdahl and Jackendoff (1983), but we did not investigate it yet. We can require that f preserves weights in a sense we now make precise.

Let us begin with Theory A, which is simpler. Since every note is interpreted, we can associate every vertex $x \in V$ with a vertex $g(x) \in V'$ such that $f(g(x)) = x$ (g is the reverse function of f). We then require, for M to be true of S, that $\forall x, y \in V,\ p(x) \leqslant p(y) \Rightarrow p'(g(x)) \leqslant p'(g(y))$. This ensures that more important musical events will be interpreted as more important events in the situation.

Things are bit more complex with Theory B, where all notes need not to be interpreted (whence the function g doesn't necessarily exist). Since heads are now implemented as weighted events, we will require that weighted musical events are interpreted. This can be done by requiring that if a vertex $y \in V$ has greater weight than another vertex $x \in V$ and if x is interpreted, then y is interpreted too (and by a more weighted vertex in V'). Nevertheless, we think that this global condition is too strong and should be local: it may happen that some part of the musical snippet describes the situation in a very fine–grained fashion (*i.e.* almost every note is interpreted) while another does it in a more coarse–grained fashion, and yet the former's notes do not have bigger weights than the latter's. Local conditions in trees have been formalised in linguistics through the notion of c–command[15]. Formally, our localised condition becomes: $\forall y \in V,\ (\exists x \in \mathrm{Im} f \text{ s.t. } y \text{ c–commands } x \text{ and } p(y) \geqslant p(x)) \Rightarrow y \in \mathrm{Im} f$. We also add the weight–preservation condition as in the other theory.

An alternative notion to weight would be that of reduction[16]. Intuitively, a musical passage can be reduced to another if it is heard as an elaboration of it — *e.g.* an ornamented version. The point is that if the tree–structure of a musical snippet encodes its reduction steps — as it

[14] though the same intuition is behind

[15] Abbreviation for constituent–command. Here we say that a node x c–commands a node y if it dominates it or is a sister of one of y's ancestors.

[16] At least two notions of reduction are discussed in Lerdahl and Jackendoff (1983): time–span reduction and prolongational reduction. For more details, see Chapter 5 and following.

is the case of the structures that Lerdahl and Jackendoff (1983) deal with[17] — weight functions may be redundant because it seems that low–weighted musical events are those which disappear after a few reduction steps. One advantage of replacing weight by reduction steps is that it now comes along with the structure and does not need to be computed separately. One drawback is that we lose the local character of weight functions, because every local branch will now reduce at the same speed[18]. We leave a closer investigation of these theoretical possibilities of implementing heads for future research.

4.3 The meaning of uninterpreted notes

Assuming now that some notes remain uninterpreted, as in Theory B, what would be their semantic role? Let us give a few clues.

We saw in the case of Figure 4 that the added notes change the *character* of the walking–situation, but do not seem to add extra events. There could be several variants of such a phenomenon, regarding how these extra notes affect the semantics of the whole.

According to one variant[19], each extra note modifies the meaning of its local branch. For instance it could be that the relevant reduction level in the case of Figure 4 is the beat level, so that each beat is viewed as a semantic atom, packaging all the information of its musical sub–events. The second and fourth beats of each bar thus have a semantics computed from the two quavers it is made of, and this may indicate that it is, for example, a particularly supple bounce. That is, each bounce is further characterised by this extra note. According to an alternative variant, extra notes are first ignored, leading to the same denotation set as with the simple snippet from Figure 1. Only in a second stage will the extra notes add the inference that the walking–situation is more bounced or more energetic; or it will give the listener clues about the mental state of the walking character (according to our informants, he is cheerful and happy, or even wanting to dance). Using a (possibly dubious) analogy from language, the former variant predicts that low–weighted notes behave more like adjectives modifying a noun phrase (here: a musical event), while the latter predicts they behave more like adverbs, modifying the whole sentence. As we have no clue for favoring one over the other — leaving aside the fact that both could partly hold together — we will not say more than just the discussion of this example, and leave these issues for future research.

4.4 Is every event musically represented?

As a reverse question, we may ask if every event in the situation should be represented by a note in M. This seems to be hardly the case, since a situation can never be exhaustively described even with language.[20] This seems to invalidate our theory, since we required that the situation tree is 'contained' in the musical tree, which is much smaller. However, we can get out of this problem by considering *reductions* of the situation tree[21] or by saying that we will assimilate the situation with what is relevant in its perception by a given agent. We will then ask that

[17] The trees here are not directed rooted trees as generally understood in mathematics, but rather something equivalent to directed rooted trees with heads, though a bit more complex.

[18] Typically, with time–span reduction, each group of two quavers will be at some point replaced by a single crotchet. But it could well be that in some branch each quaver is interpreted, while in some other only one of them is.

[19] Thanks to Philippe Schlenker for suggesting this variant.

[20] Think, for example, of the whole subtlety in the gestures of a character, or of all the particle interactions that take place everywhere.

[21] Just like there are notions of reduction for music, we can posit that the same things exist with situations, as suggested in Jackendoff (2009).

every head event in the situation is represented by a head event in the musical snippet.

All this being said, let us state as a summary a final formal formulation of one variant of Theory B. Theory A could be straightforwardly formulated in a very similar way.

Theory B. *Let* $M = (V, E, p)$ *be a musical snippet and* S *a situation.* $M \models S$ *if, and only if,* $M \models S$ *in Schlenker's sense, and moreover there exists a reduction* $S' = (V', E', p')$ *of* S*, and an injective root–preserving function* $f : V' \to V$ *such that:*

1. $\forall x, y \in V',\ (x, y) \in E' \Rightarrow f(x) \sim f(y)$
2. $\forall x, y \in V',\ p'(x) \leqslant p'(y) \Rightarrow p(f(x)) \leqslant p(f(y))$
3. $\forall y \in V,\ (\exists x \in Imf \text{ s.t. } y \text{ c–commands } x \text{ and } p(y) \geqslant p(x)) \Rightarrow y \in Imf$

5 Conclusion

The investigation of walking–situations and walk-denoting music enlightened the necessity to come up with a theory of musical events. This theory needs to provide rules involving the structural role of each musical event determined through the rigorous analysis of musical structures, and explain how this impacts the very possibility for each event to be interpreted, *i.e.* to have a counterpart in the denotation. Besides, our goal was to account for how this interpretive semantic mechanism works (*i.e.* what happens to the musical structure when interpreted, and how the formal tree–like structure of the music is related to the formal tree–like structure of the events it denotes). Further research will investigate the experimental extensions of this theoretical work, in order to check whether it has some cognitive reality in listeners.

References

Daniel Harasim, Martin Rohrmeier, and Timothy J. O'Donnell. A generalized parsing framework for generative models of harmonic syntax. In *Proceedings of the 19th International Society for Music Information Retrieval Conference, ISMIR 2018, Paris, France, September 23-27, 2018*, pages 152–159, 2018.

Ray Jackendoff. Parallels and nonparallels between language and music. 26(3):195–204, 2009. ISSN 0730-7829. doi: 10.1525/mp.2009.26.3.195.

Fred Lerdahl and Ray Jackendoff. *A generative theory of tonal music.* The MIT Press, Cambridge. MA, 1983. ISBN 0262120941.

Philippe Schlenker. Outline of music semantics. 35(1):3–37, a. ISSN 0730-7829, 1533-8312. doi: 10.1525/mp.2017.35.1.3. URL http://mp.ucpress.edu/lookup/doi/10.1525/mp.2017.35.1.3.

Philippe Schlenker. Prolegomena to music semantics. 10(1):35–111, b. ISSN 1878-5158, 1878-5166. doi: 10.1007/s13164-018-0384-5. URL http://link.springer.com/10.1007/s13164-018-0384-5.

Jeffrey Zacks, Barbara Tversky, and Gowri Iyer. Perceiving, remembering, and communicating structure in events. *Journal of experimental psychology. General*, 130:29–58, 04 2001. doi: 10.1037//0096-3445.130.1.29.