



HAL
open science

Investigating musical meaning: Empirical evidence for a music semantics

Léo Migotti Ramponi

► To cite this version:

Léo Migotti Ramponi. Investigating musical meaning: Empirical evidence for a music semantics. Cognitive science. Ecole Normale Supérieure (ENS), Paris, FRA., 2023. English. NNT: . tel-04294445

HAL Id: tel-04294445

<https://theses.hal.science/tel-04294445v1>

Submitted on 19 Nov 2023

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution 4.0 International License



THÈSE DE DOCTORAT
DE L'UNIVERSITÉ PSL

Préparée à l'École Normale Supérieure

**Investigating musical meaning:
Empirical evidence for a music semantics**

En vue d'une soutenance par

Léo Migotti

Le 31 août 2023

École doctorale n°158

**Cerveau, Cognition,
Comportement**

Spécialité

Sciences Cognitives

Préparée à

Institut Jean Nicod
(CNRS - ENS - EHESS)

Composition du jury :

Emmanuel CHEMLA *Examineur*
LCSP, Ecole Normale Supérieure

Claire PELOFI *Examinatrice*
New York University

Jonah KATZ *Rapporteur*
West Virginia University

Isabelle CHARNAVEL *Rapporteuse*
Université de Genève

Philippe SCHLENKER *Directeur de thèse*
Institut Jean Nicod (CNRS-ENS-EHESS)
New York University

Jean-Julien AUCOUTURIER *Co-encadrant*
CNRS, FEMTO-ST

To my grandfather Luigi, who used to sing
Elle était belle et souriait.
I'm sure you're still singing where you are now.

Acknowledgements

First and foremost, I am extremely grateful to my main supervisor Philippe Schlenker for his invaluable advice and continuous support during my master and PhD study. His knowledge and experience have not only been a constant inspiration, but they have helped me grow as a better scientist and a better person.

I would like to express my sincere gratitude to my co-advisor Jean-Julien Aucouturier, whose skills and presence have allowed me to broaden my perspectives on music cognition and engage in crucial critical thinking on my work.

I am deeply indebted to Emmanuel Chemla, who has not only provided me with extremely insightful and continuous feedback on my work, but has also greatly helped me raise my awareness of experimental procedures and sharpen my skills in experimental design and statistical analyses.

Many thanks to Claire Pelofi for her astute feedback and enlightening discussions during my PhD committees, and to Isabelle Charnavel and Jonah Katz who accepted to review this dissertation.

Collaborations have also been essential to this PhD project. I want to express my deepest gratitude to my esteemed colleague and friend Janek Guerrini and his unwavering kindness and trust. I am also extremely grateful to Mélissa Berthet, who introduced me to the wonders of ethology and made the first half of my PhD a both thrilling and peaceful journey. Many thanks too to Léo Zaradkzi, who I had the great pleasure to collaborate with in the early stages of my PhD. Special thanks to Alessandro Ansani, with whom I have had the chance to work, for his humanity and unfailing understanding. I would also like to sincerely thank Jeremy Kuhn and Lyn Tieu for the time spent on ongoing collaborations. Thank you to all members of the Institut Jean Nicod community and the Cognitive Science Department at ENS.

Travels have also made this experience incredibly inspiring. I would like to express my gratitude to all people who contributed to making those highly stimulating moments possible. In particular, thank you to all the RITMO team at the University of Oslo for a wonderful and unforgettable experience. I am also greatly thankful to Pritty Patel-Grosz for her help in the process, and for very enriching discussions and guidance from the very

first moments of this adventure.

These past two years have also been a great opportunity for me to teach. I would like to thank Denis Bonnay, Benjamin Simmenauer and Françoise Sackrider for trusting me in teaching at the Institut Français de la Mode. This teaching experience has already been determining for my working life.

I could not have undertaken this journey without all the extraordinary support I have received from family and friends, and words can hardly express how thankful I am (but maybe music could do it; and I know each of you is associated with — at least — one song in my mind). Thank you from the bottom of my heart to my mother, my sister and my father for their unconditional love and support. None of this could have happened without them. I am so very lucky to have received inexorable attention and support from my amazing friends, whom I will thank personally.

The past few years have also given me the opportunity to reconnect with art, and in particular with writing and musical practice. This would not have been possible without the very talented artists from the theatre company Le Printemps du Machiniste. Thank you so very much for their trust. I am now certain that art and science do not only connect in the world but also in my heart.

Thank you to anyone whom I might have forgotten and who contributed in one way or another to this unique experience. May music be with you!

Résumé

La musique a-t-elle un sens ? Et si oui, quel type de sens ? Dans cette thèse, nous présentons des données expérimentales qui soutiennent l'idée que la musique possède une sémantique : (i) la musique est capable de faire référence à une réalité extra-musicale, et (ii) elle le fait selon des règles systématiques. Afin de tester ces deux hypothèses, nous présentons quatre paradigmes expérimentaux dans lesquels la musique interagit avec d'autres facultés cognitives telles que le langage, la perception visuelle ou le mouvement corporel.

Dans la première expérience, nous avons présenté aux participants des phrases hybrides dans lesquelles des stimuli musicaux remplaçaient des mots pour évaluer le sens de ces phrases. Les résultats montrent que le contenu informationnel de la musique se répartit selon les différentes catégories de la typologie linguistique (c'est-à-dire les différents types d'inférences déclenchées par les phrases complexes) : les mécanismes cognitifs classifiant le sens dans le langage peuvent donc traiter l'information musicale.

Dans la deuxième expérience, les participants devaient juger des scènes audiovisuelles dans lesquelles des propriétés musicales telles que la hauteur des notes, les nuances et le timbre étaient manipulées et associées à des objets visuels. Les résultats montrent que (i) la hauteur des notes et les nuances peuvent être interprétées comme le niveau d'énergie ou la distance d'un objet visuel, (ii) ces propriétés peuvent être utilisées pour identifier un objet lorsque plusieurs objets sont présents, et (iii) le timbre est interprété comme la nature même d'un objet. Ces résultats montrent que les propriétés musicales ne sont pas seulement associées à des scènes générales, mais qu'elles sont liées à des objets contenus dans ces scènes.

La troisième expérience est une étude de cas sur la musique qui évoque la marche. Nous avons cherché à déterminer les conditions de cette association grâce à une tâche de préférence dans laquelle les participants jugeait des stimuli selon leur capacité à faire référence à un personnage en train de marcher. Les résultats montrent qu'au moins cinq propriétés sont impliquées dans cette association : trois d'entre elles (stabilité, alternance, binarité) sont liées aux propriétés structurelles d'une marche, et deux d'entre elles (consonance absolue et proximité relative des accords) sont liées aux propriétés physiques d'une marche. La consonance est interprétée comme la stabilité de chaque pas, tandis que la proximité des accords est associée à la symétrie de la démarche. Ce modèle rend

par ailleurs mieux compte de la sémantique de la musique évoquant des marches que les modèles précédents et fournit des données préliminaires suggérant que chaque événement musical est interprété comme un événement du monde réel.

La quatrième expérience propose enfin de tester la possibilité de trouver des traces de la sémantique de la hauteur des notes sur le mouvement corporel. Les participants avaient pour tâche de marcher en rythme avec une série de sons dont la hauteur variait aléatoirement. Nous avons constaté que leurs pas étaient à la fois plus longs et réalisés avec une plus grande force sur les notes basses. La similarité entre l'effet physique produit par une note plus basse (associée à davantage de poids) et la sémantique d'une note plus basse (associée à des objets plus gros ou ayant moins d'énergie) suggère qu'il existe une interaction entre les représentations musicales et le système moteur.

Dans l'ensemble, nos résultats fournissent des preuves en faveur d'une sémantique de la musique : la musique peut faire référence à une réalité extra-musicale selon un ensemble de règles systématiques. Ils montrent également que la musique interagit avec d'autres systèmes cognitifs, qu'ils portent directement un sens (langage) ou non (mouvement corporel), ce qui soulève des questions sur la manière dont les effets induits par la musique se rapportent à la signification musicale.

Mots clés : musique, sémantique, sens, cognition musicale

Abstract

Does music have meaning? And if so, what kind of meaning? In this dissertation, we provide experimental data supporting the view that music has a semantics: (i) music is able to refer to an extra-musical reality, and (ii) it does so in a rule-governed fashion. We present four experimental paradigms designed to test these two assumptions by having music interact with other cognitive faculties such as language, visual perception, or body motion.

In the first experiment, we presented participants with hybrid sentences containing music in lieu of words and tried to determine what meaning they get from such sentences. The results show that the informational content of embedded musical stimuli could be divided among the different slots of the linguistic typology (i.e. the different types of inference that complex sentences trigger). This suggests that the cognitive mechanisms classifying meaning in language can accommodate musical information.

In the second experiment, we had participants judge audiovisual scenes in which musical properties such as pitch height, dynamics, and timbre were manipulated and paired with visual objects. The results show that (i) both pitch height and dynamics can be semantically interpreted as the energy level or the distance of a visual object, (ii) pitch height and dynamics can be used to retrieve which object is being referred to when multiple objects are present, and (iii) timbre is interpreted as the very nature of an object, hence the timbre/object association cannot be violated in audiovisual scenes. Together, these results show that musical properties are not only associated with general scenes, but that they are bound to objects in those scenes.

In the third experiment, we built a case study about music that evokes walking. We sought to determine what it takes for a musical piece to evoke a walk through a preference task in which participants had to pick stimuli based on how well they evoked a character walking. The results show that at least five properties are involved in the association of a musical stimulus with a walk: three of which (steadiness, alternation, binarity) relate to structural properties of a walk, and two of which (absolute consonance and relative chord proximity) relate to physical properties of a walk such as stability. Consonance was arguably interpreted as the stability of each step, while chord proximity was associated with gait symmetry. The results further show that this fine-grained model accounts for the semantics of music evoking walks better than previous models and provide preliminary

evidence that each musical event might have to be interpreted as a real-world event.

In the fourth experiment, we finally checked whether it is possible to find footprints of the semantics of pitch on body motion. Participants were tasked to walk in synchrony with a series of sounds varied in pitch. We found that their steps were both longer and produced with greater force on lower pitches. The similarity between the physical effect produced by lower pitch (mimicking an increase in weight) and the semantics of lower pitch (associated with bigger objects, or less energy) suggests that there is an interaction between musical representations and the motor system.

Together, our results provide evidence for a music semantics: music can refer to an extra-musical reality according to a set of systematic rules. They also show that music interacts with other cognitive systems either directly bearing meaning (language) or not (body motion), which raises questions as to how music-induced effects relate to musical meaning.

Keywords : music, semantics, meaning, music cognition

Contents

Acknowledgements	ii
Résumé	iv
Abstract	vi
Table of Contents	vii
1 General introduction	1
2 Meaning from music embedded in language	33
3 Meaning from music in audiovisual scenes	83
4 Meaning from music alone	113
5 Music-induced effects on body motion	167
6 General discussion and conclusions	188
Publications	205

Chapter 1

General introduction

Contents

1	Musical structure without meaning?	4
1.1	Musical syntax	5
1.2	No meaning	6
2	Musical meaning(s): a typology	6
2.1	An internal meaning	7
2.1.1	Musical meaning as musical form	7
2.1.2	Musical meaning as expectations	8
2.2	A hybrid meaning	8
2.2.1	Musical meaning as emotions	8
2.2.2	Musical meaning as intention attribution	11
2.2.3	Musical meaning as musical forces	12
2.3	An external meaning	13
2.3.1	Musical meaning as social uses	13
2.3.2	Musical meaning as information about objects	14
2.3.3	Musical meaning as inferences about the world	15
3	Music semantics	15
3.1	Musical truth	16
3.2	Musical inferences	18
4	Outline	19
4.1	Meaning from music embedded in language	20
4.2	Meaning from music interacting with visual scenes	22
4.3	Musical meaning from music alone	24
4.4	Music-induced body effects	26
5	Bibliography	29

I consider that music is, by its very nature, essentially powerless to express anything at all, whether a feeling, an attitude of mind, or psychological mood, a phenomenon of nature... Expression has never been an inherent property of music. That is by no means the purpose of its existence.

Igor Stravinsky (1935)

It is very interesting that a composer such as Igor Stravinsky, famously known for having composed some extraordinarily expressive masterpieces such as *The Rite of Spring* (1913), argues that music ‘is powerless to express anything’. From the first bars of the piece indeed,¹ the famous musical motive to be repeated, developed and enriched with many innovative orchestration techniques, seems to immediately convey a sense of dialog between some objects, animals, insects, or natural phenomena. Stravinsky himself described the introduction of the piece as ‘a swarm of spring pipes.’ When it premiered in 1913 at the Paris Champs-Élysées theater, the *Rite of Spring*, which Stravinsky composed to accompany a ballet choreographed by Russian choreographer Vaslav Nijinsky, was both acclaimed as a major revolutionary piece and described as scandalous and outrageous. The piece was accused of relying too highly on rhythm, unusual stress patterns, and dissonances; to a much greater extent than had ever been done before.

The year before, in 1912, Vaslav Nijinsky had already caused a scandal when performing on Debussy’s *Prélude à l’après-midi d’un faune*.² The piece, which premiered in 1896 and was later choreographed, had also been criticized for almost opposite reasons, and in particular the lack of regular beat or rhythmically salient patterns. Inspired by a poem by French poet Stéphane Mallarmé telling the story of a lovesick faun contemplating nature, the piece is also extremely evocative of natural sounds and has since been acclaimed for its unique pastoral atmosphere. Although very different in the rest of the pieces — but with similar disagreement from the audience — the two pieces share very similar initial parts, both of which involve almost exclusively wind instruments and soft recurring melodic patterns, with which both composers managed to evoke nature. Later, Debussy writes:

¹See for instance [this performance](#) without choreography by the London Symphony Orchestra, conducted by Sir Simon Rattle (2017).

²See [this performance](#) without choreography by the London Symphony Orchestra, conducted by François-Xavier Roth (2017).

There is nothing more musical than a sunset. He who feels what he sees will find no more beautiful example of development in all that book which, alas, musicians read but too little - the book of Nature.

Stravinsky would have probably not agreed with the idea that an event such as a sunset could have anything musical to it; or at least, he would not have agreed that music could express anything about it, let alone represent it. Somehow, Debussy takes a different stance on the ability of music to depict natural events, and many of his other pieces, such as *La Mer* (1905),³ were similarly evocative of natural things or even mimicking natural sounds.

Together, these two quotes are just an instance of the ever-lasting debate on whether music can convey information about the world, which is very much related to the question of whether music can be said to have meaning. Many divergent claims have been given on this matter throughout the history of music theory and, later, music psychology and music cognition. We will now illustrate two opposite claims: first, that music does not have meaning (at all); second, that music does have meaning (we will then specify what this musical meaning can be).

1 Musical structure without meaning?

We will first explore the idea that music does not have meaning, and we will need to introduce a distinction between musical *meaning*, which refers to what music is potentially about, and musical *syntax*, which refers to principles governing musical structure and form. Crucially, a system that has syntax does not necessarily have meaning too: it is possible to create an infinity of languages that obey rules of well-formedness which do not have any meaning.⁴

³See for instance [this performance](#) by the French National Orchestra with conductor Cristian Măcelaru.

⁴An example using meaningless syllables is given in Schlenker (2017)

1.1 Musical syntax

It is today uncontroversial that music has a syntax, defined as a set of rules that determine whether a musical piece is well-formed.⁵ The pioneer work by Lerdahl and Jackendoff (1983) established several categories of such rules for tonal music, and on top of these well-formedness principles, preference rules that account for why and how listeners assign certain mental structures to music instead of other possible equally well-formed ones. More recent accounts have updated this view (Rohrmeier, 2011) and applied rules from other musical traditions such as jazz.⁶ But crucially, no investigation of musical form and the systematic rules that music must satisfy to be considered well-formed has to commit to whether music has meaning or not. This also holds (i) regardless of how similar to the syntax of natural languages musical syntax is, and (ii) despite the fact that natural languages have both. In a paper investigating the similarities between the formal properties of language and music, Katz and Pesetsky (2011) provided arguments in favor of the ‘Identity thesis’ for music and language, which states that “All formal differences between language and music are a consequence of differences in their fundamental building blocks. In all other respects, language and music are identical.” In other words, what makes music different from language is meaning-irrelevant: the only reason for which the two systems differ is because their constituent parts are not the same, but if we put that aside, then the formal structure of both systems is identical. This identity is therefore inherently syntactical; hence one cannot predict from this analogy that music should have meaning. We will now continue with an exploration of the hypothesis that, despite having a syntax, music does not have meaning.

⁵Note that, as argued in Schlenker (2017), we do not need to either take a stance on how this syntax works to support the claim that music has meaning; although we have good reasons to believe that musical syntax interacts with what we will soon call music semantics.

⁶See for instance Granroth-Wilding and Steedman (2014) on a model accounting for how chord sequences are mapped onto mental representations about the hierarchical and structured relation between chords in listeners.

1.2 No meaning

Many arguments have been raised to support the view that music does not have meaning. But a prominent one has been targeted at the plurality of things that music evoke to make the point: in short, music cannot have meaning because different listeners assign different meanings to a same musical piece. This argument was famously reused by American conductor and composer Leonard Bernstein (2005), using an example from Strauss's Variation II from symphonic poem *Don Quixote*. The reason for which music does not have meaning, Bernstein argued, was that one could not spontaneously know what that piece was supposed to represent without knowing the program. It was then possible to come up with a 'wrong' story which the piece could just as much represent, and Bernstein came up with a story about Superman to illustrate that the intended meaning of the piece could be otherwise. But this argument might actually not support the view it intends to. The very fact that multiple interpretations are possible, as noted by Schlenker (2022), perfectly illustrates the opposite claim, for the 'wrong' story was actually almost perfectly isomorphic to (i.e. had the very same structure as) the original plot. Far from showing that music has no meaning, this argument therefore gives a clue about the nature of musical meaning. Not only does music have meaning, but it has an *abstract* meaning, in the sense that multiple scenes can be represented by a same music, as long as they share some of their properties. Just as in language, in which sentences are true of different situations, music can be *true of* different situations.⁷

2 Musical meaning(s): a typology

Arguing that music can be true of situations already commits us to a very specific theory of musical meaning, while many authors have given arguments supporting the claim that music has meaning without necessarily appealing to the concept of truth. First, it has been argued that music only has an internal meaning, i.e. that musical meaning lies in the form of music itself. By contrast, other theories have argued that music has an extra-

⁷We come back to this notion of musical truth a bit later.

musical meaning,⁸ i.e. that music is associated with things in the extra-musical world, either real or imagined. Hybrid theories have posited that musical meaning is reducible to emotions; we will discuss whether this rather amounts to an internal of extra-musical meaning in the following sections.

2.1 An internal meaning

2.1.1 Musical meaning as musical form

Among proponents of internal musical meaning, Austrian music theorist Hanslick (1957) defined music as ‘successions and forms of sound’, and argued that ‘these alone constitute the subject’ (of music). In other words, the only subject matter of music is precisely what it is made of, its own components: in short, its form. It is of course uncontroversial that music conveys information about its own form; and it is also right that this does not entail at all that it should be able to convey information about something else. While Hanslick (1957) recognizes that music triggers feelings and emotions (as is today commonly admitted (Blumstein et al., 2012; Juslin and Laukka, 2003; Koelsch, 2012)), he stresses that music is however unable to represent these emotions. Under this view, a piece of music can for instance make someone happy, but it cannot represent a happy scene, or a scene in which someone is happy. But that listeners do not get meaning from music is not incompatible, Hanslick argued, with their active engagement in listening. For him, ‘two listening modes’ are available to the listener: a passive one, comparable to just hearing music without actively engaging in music; and an active one, described as a state in which the listener attempts to follow the composer’s intention and anticipate what should come next in a piece. Interestingly, the existence of this second active mode seems to be an argument in favor of a music *pragmatics*, i.e. reasoning patterns involving predictions about what is to come next in a musical piece, and updates on these predictions. These reasoning patterns however remain ‘internal’ in the sense that they do not involve anything from outside music: trying to predict what is to come next in a musical piece is a reasoning still targeted at musical form. We explore another theory

⁸We here reuse this term coined by Koelsch (2012).

of musical expectations which emphasizes the ‘external’ emotional counterpart of these prediction mechanisms in the following section.

2.1.2 Musical meaning as expectations

If our understanding of music solely relies on predictions on its form, it follows that grounding musical meaning in expectations amounts to considering that musical meaning lies within music. But expectations on musical form have also been argued to trigger emotions in listeners. Huron (2006) notably developed a theory in which musical predictions are systematically paired with ‘affective responses’, i.e. emotions. In this theory, two types of response are associated with anticipations of musical events before they occur: imaginative and tension responses respectively trigger emotions about how one feels about what they think is gonna happen next, and about how ready one feels based on how a musical event is being prepared. Prediction, reaction and appraisal responses follow once a musical event has occurred: they encompass the satisfaction or dissatisfaction a listener feels based on whether the expectations were satisfied, with one’s assessment of one’s own reaction to the musical event (such as surprise), and a more general appreciation of how things have turned out upon reflection. But Huron does not, per se, take a stance on whether these different mechanisms are what constitute musical meaning; yet, such theories have been developed.

2.2 A hybrid meaning

In this section, we discuss the theoretical implications of considering that musical meaning reduces to music-induced emotions.

2.2.1 Musical meaning as emotions

Meyer (1956) notoriously argued that music has meaning and that this meaning could be reduced to the emotions music triggers, which themselves originate in whether musical expectations are satisfied or not. For Meyer, there are two sources of musical expectations: (i) universal perceptual principles and (ii) ‘archetypes’ (namely, expectations from

specific musical conventions governing musical systems which listeners are familiar with).⁹ Consistent with Huron (2006), Meyer claimed that musical emotions are triggered by musical expectations, but his theory crucially differs in stating that the meaning of music is reducible to these emotions. Note that this is already different from saying that musical meaning solely lies in the information it conveys about its own form and its constituent blocks, for if it can trigger musical emotions, it is already in relation to something that is extra-musical. If (i) music triggers emotions, and if (ii) these emotions constitute musical meaning, then given that (iii) emotions are external to music, it follows that musical meaning is external to music.

Yet for Meyer, music “operates as a closed system, that is, it employs no signs or symbols referring to the non-musical world of objects, concepts, and human desires.” An analogous claim was later made by Davies (1994), for whom music neither had the ability to describe nor to depict: the existence of musical emotions says nothing about music as referring to non-musical things.¹⁰ But Meyer also stressed that it was “necessary to emphasize that the prominence given to this aspect of musical meaning [i.e. emotions] does not imply that other kinds of meaning do not exist or are not important”, and this later claim leaves open the question of the existence of an extra-musical *referential* meaning besides one that strictly consists of the set of emotions generated by musical expectations.¹¹ One might however argue that generating emotions entails representing emotions. In Schlenker (2017) for instance, music is argued to convey emotional information through two different mechanisms. First, it is possible that a listener represents emotions of non-musical objects: in this case, a musical piece does not necessarily make the listener feel sad at first, but they understand that the music is referring to something

⁹Note that this distinction is close to that introduced by Schlenker (2017) on the distinction between meaning-bearing musical properties lifted from normal auditory cognition (such as loudness, plausibly universal) and those that are specific to music (such that harmonic rules) and therefore probably those that are specific to particular musical systems such as the tonal system as well.

¹⁰Unless, as we will see, one has a theory of how musical emotions relate to the extra-musical world, and in particular to emotions of music-external events.

¹¹Here and after, we will use the term ‘referential’ (also known as ‘denotative’) as commonly used in linguistics to characterize the property of an item with a given form (be it linguistic - such as a word, sign, sentence - or non-linguistic - such as a gesture, a picture, a visual representation, a sound or a musical sequence) to have the external world as referent, and not any part of the form or its properties itself. Referential meaning typically does not include connotations, cultural associations, or pragmatic enrichments.

sad (a sad character, for instance), and this can *then* affect the emotional state of the listener ('by contagion'). Second, it is possible that music is interpreted as events *experienced* by listeners themselves. Under this view, music does refer to extra-musical events, but those events are *internal* to the listener. This is crucially different from the internal accounts we have mentioned such as Hanslick's, where meaning lies within musical form; here, meaning lies within the listener's experience of music. Hence, the existence of musical emotions is compatible with music referring to an extra-musical reality, but this is not what most accounts of musical emotions such as Meyer's presuppose.

Aside from the structural similarities between music and language sketched in the previous section, several authors have more specifically argued that music conveys meaning because it shares some of its fundamental constituent properties with other communicative systems such as human speech or animal vocalizations that are known to convey emotions. Ilie and Thompson (2006) demonstrated that pitch height, intensity and rate have effects on valence judgments, energy arousal and tension arousal in both speech and music. Along the same line of work, Bowling et al. (2010) found correlations in the spectra of intervals and speech: major intervals have a spectrum that is similar to that of excited speech, while the spectrum of minor intervals is more similar to that of subdued speech. Bowling et al. (2012) reported similar results when comparing Western to South Indian music, indicating that the music/speech analogies are neither exclusive to a specific language nor to a specific musical tradition. Together, these results suggest a universal code for conveying emotional information, a code which might even be shared more broadly, beyond music and speech. Juslin and Laukka (2003) found analogous similarities between vocal performance and instrumental music, and Bedoya et al. (2021) showed that applying acoustic modifications modeled after human voice to instrumental sounds maintained emotion recognition. And beyond audition, Sievers et al. (2013, 2019) has provided evidence that this emotional code is also shared across modalities (in particular between audition and vision). It has finally been argued that this code or part of it is not only shared across domains but across species as well. For instance, Blumstein et al. (2012) found that non-linearities (sudden rises or drops in frequency), that signal emo-

tional states in the vocalization systems of many vertebrates, do convey similar effects in humans.

In sum, it is well established that music conveys emotions, be it because it reuses some features already present in language or vocalization that do convey emotions, or for other reasons. While Meyer (1956) states that these emotions are what constitute musical meaning, the other references we provided do not necessarily commit to this definition. In any event, under the assumption that musical meaning can be reduced to music-induced emotions, then we tend towards a notion of musical meaning that is directed outward music. But again, this is different from saying that music can represent emotions — unless one has an explicit theory of how these emotions relate to emotions experienced by extra-musical objects, or by listeners themselves.

2.2.2 Musical meaning as intention attribution

Some research has evidenced the possibility for musical components such as notes or chords to be associated with imagined fictional agents having intentions, regardless of musical emotions (Maus, 1988). These theories echo the famous animation by Heider and Simmel (1944) (accessible [here](#)) in which geometrical shapes are very easily assigned intentions. Just as such abstract objects can be assigned intentions by manipulating some properties of their movements (in particular by creating behaviors which physical laws alone could not account for), these theories assign intentions to musical events associated with different ‘fictional agent types.’¹² Maus (1988) claims that listeners associate musical works with such imagined agents, which serve as ‘repositories’ for psychological states. Crucially, these associations are indeterminate and ad hoc since they vary from listener to listener and cannot be predicted. This may be a problem for these theories, which seem under-specified as far as the nature of these imaginary agents is concerned. While under a certain interpretation, musical events (or musical ‘forms’) are associated with agents that might fall beyond the scope of the piece itself, Maus (1988) also claims that some agents

¹²Despite the conceptual similarity, in Heider and Simmel (1944), it is the geometrical shapes themselves which are assigned intentions; while in Maus (1988), musical events (or musical ‘forms’) are associated with agents that might fall beyond the scope of the piece itself.

are ‘coextensive with the musical work’ (i.e. are part of the piece itself). Monahan (2013) argues that this account resonates with that of Cone (1982), which states that ‘lower-level agencies are understood collectively as components of a higher-order persona that is coextensive with the work as a whole.’ Under this view, the aggregation of individual agents results in the abstract attribution of an intention (or even a ‘mind’) to the piece itself. Here again, this composite intention seems to lie within the scope of the musical piece. Yet Maus (1988) also allows for these intentional agents to be ‘fictionalized versions of the composer or the performers’: the idea here is that listeners attempt to reconstruct a performer’s or a composer’s intention and interpret music relative to their viewpoint. Although this part of the theory is very similar to that of Hanslick (1957) presented earlier, the under-specification of the nature of these fictional agents makes the general theory hybrid, in that it allows for musical meaning to refer to agents understood as being part of the piece, or to refer to external agents. In sum, these theories seem to commit neither to a pure notion of internal meaning, nor to one of external meaning.

2.2.3 Musical meaning as musical forces

Another line of research has explored how musical properties are interpreted in terms of auditory counterparts of physical forces such as gravity, magnetism, and inertia. For instance, Larson (2012) came up with a model in which pitches entertain similar parallel between themselves, explaining how listeners experience meaning as reducible to motions constrained by musical forces throughout the course of a piece. Honing (2003) investigated how physical models of slowing down motion can account for how musical pieces generally end with a final *ritardando* (i.e. a decrease in tempo marking the end of the piece). Although the application of physical models to interpretative musical facts does appear to create constraints on which interpretations (in terms of performance, not in semantic terms) are licensed, the authors call for caution in applying these methods systematically. Besides, that the rate at which speed decreases in a real world slowing-down motion matches that of the rate at which music speed decreases at the end of a piece can in theory be explained in purely syntactical terms, if, for instance, a performer tries align

the musical structure with that of the event. And in any event, it seems here that these abstract forces merely refer to the physical work by analogy.

2.3 An external meaning

Although some of the previous accounts of musical meaning already contain a sense of reference to non-musical things such as emotions, imaginary agents, characters, or forces, they can all be considered hybrid in that they do not need to take a strong stance on whether musical meaning lies within or outside music itself. In this section, we finally explore different accounts which less ambiguously claim that musical meaning exists in virtue of its connection to the outside world.

2.3.1 Musical meaning as social uses

Music can first connect to the outside world in virtue of the numerous actual events and social contexts in which it is used, a view which led Cross and Woodruff (2009) to argue that “Music and language are both part of the human communication toolkit.” Under the assumption that music and language have co-evolved from an ancestral form, music is considered to be an optimal signal which is able to “manage situations of social uncertainty.” While music is often considered as a communicative system which primarily aims at communicating emotions, it seems to convey other kinds of information. Music is notably widely found to play a crucial role in regulating social interactions through social bonding when the cohesion of a group is somehow threatened, either from outside agents or from within. Music accompanies major life transitions in many societies (Blacking, 1976; Schulte-Tenckhoff and Feld, 1988), and has also been widely used to create or recreate a sense of community in threatened or minority groups (Slobin, 1993). Music also plays a crucial role in emotional self-regulation (Juslin and Sloboda, 2001), and is almost universally associated to infant care (Trevvarthen, 1999). But what kind of information does music provide in these contexts, exactly? Here, it seems that musical meaning is defined as *social use*, a view which has been given some more recent empirical evidence in Mehr et al. (2019), which show that most musical features in songs are cross-

culturally associated with properties of the contexts in which the songs are performed, and for which a typology can be given based on three dimensions (formality, arousal, and religiosity).

The correlations between some music types and social uses do however not quite entail that music can refer to extra-musical things. It could indeed be that those associations are mere conventions, and that a given musical piece in a given situation does not per se represent that situation, and that is just empirically associated with it. However, this need not be incompatible with the view that music has referential meaning. One might indeed argue that although partly conventional, musical conventions (in terms of social use, not in terms of composition) are established because they evoke mental representations that are consistent with the context in which a given musical piece is played or sung. For instance, lullabies might have spread in the context of infant care not only by convention, but because they could be associated to such situations a priori.

2.3.2 Musical meaning as information about objects

Among all previous accounts, none of them (music referring to social contexts, music triggering emotions, possibly relying on shared mechanisms with speech) requires that music be able to make reference to extra-musical things. We will now turn to theories and empirical evidence which support this view more clearly.

A first well-documented line of work focuses on relations music entertains with representations of movement. Music has been argued to be associated with spatial and motion properties of virtual spaces (Clarke, 2001), or to motion properties of musical events themselves. For instance, Saslaw (1996) suggests that musical objects themselves are conceived as moving through space, while having ‘the attributes of real-world objects: weight, speed, force.’ However, these theories ambiguously stand by the fact that musical meaning is somehow still internal to music: if listeners listen to music as conveying information about musical events themselves (such as tones), it does not necessarily follow that they would assign similar attitudes to extra-musical agents. Yet, several musical properties have been experimentally shown to be responsible for associations to objects

beyond music-internal representations (Godøy and Leman, 2010). For example, Eitan (2013) explored the associations listeners make between pitch and loudness properties of music and visual objects, suggesting that loudness can be thought of as indicating distance, while pitch is notoriously mapped onto height. Musical properties such as dynamics, pitch direction, intervals, attack rate, and articulation affect motion imagery in tasks where participants have to imagine a character’s movement (Eitan and Timmers, 2010). Interestingly, while the authors note that these associations are not codified in natural language, recent work indicate that they can be iconically and visually represented in sign languages (Schlenker, 2018a).

2.3.3 Musical meaning as inferences about the world

More recent theoretical work has developed a theoretical framework in which properties of time, pitch, loudness, and harmony are integrated to account for musical inferences about virtual objects (Schlenker, 2017): ‘minimal pairs’ of musical stimuli changing only with regards to one property show that one can create contrasts in the evoked scenes. For instance, rising a musical sequence by an octave changes the size property of the denoted object: a higher-pitched musical sequence typically represents a small object better than a large one. While mainly consistent with previous research, some new insights on the rules mapping each property to its semantic interpretation are provided. This line of work is the one we will rely on the most throughout this dissertation. We discuss its theoretical details and predictions in the following section.

3 Music semantics

We have just seen that there are many arguments supporting the idea that music can refer to non-musical objects undergoing some events (such as moving), i.e. to an extra-musical reality. Yet, this could still be the case if the relation between music and what it refers to were conventional; or if it were purely subjective in the sense that it only depends on a listener’s personal interpretation. Acknowledging the existence of meaning effects triggered by music does not entail that music has a *semantics*, defined as ‘a rule-

governed way in which music can provide information (i.e. license inferences) about some music-external reality' (Schlenker, 2017). Under this view, the mechanisms responsible for the generation of mental representations from music are not arbitrary: they obey systematic rules that map musical properties to properties of an extra-musical world. This view is in particular different from already mentioned theories of ad hoc agentivity assignment such as Maus (1988), which posit that each listener parses a musical work as they please. In this section, we present the main concepts of the theoretical framework of music semantics, which this dissertation will rely on, and introduce two of its main concepts: musical truth, and musical inferences.

3.1 Musical truth

For a few years, music semantics has been developing within the broader research agenda of Super Semantics (Schlenker, 2018*b*). The core idea of this research programme is to extend the methods of formal semantics to the analysis of meaning beyond language.

Under the assumption that music has, just as language, the ability to refer to a non-musical reality, it is possible to extend the traditional definition of linguistic meaning as the set of *truth conditions* of a proposition (i.e. statements about the conditions under which a given proposition is true) to music, and to define musical meaning through truth conditions as well. The truth conditions determine under which conditions a given musical event or sequence is *true of* an extra-musical event.¹³ Now, establishing truth conditions for music requires some clarification on how music can be true of an event. A musical event M is true of an event E if that event E satisfies the relevant properties of M . In other words, a musical excerpt is true of an event if the semantic interpretation of its musical properties complies with the properties of the event relative to a set of semantic rules. This is all very abstract. Let's take a quick example to illustrate this.

We have seen in the previous section that many studies have established correlations

¹³This is a first approximation used for clarity. In Schlenker (2017), musical events are not taken to be true of events but of pairs of objects and n-tuples of events. It is also possible to define musical truth relative to situations that involve multiple events. We will for now purposely remain vague as to which of these entities music primarily refers to. The experiment presented in section 4.2 will try to address this point

between some musical properties and physical properties of objects such as size or energy.. For instance, loudness can be interpreted as providing some information about the level of energy involved in an event (Schlenker, 2017). **The louder the sound, the more energy in the event.** And here we are already: the previous statement is a semantic rule which states that (i) loudness is semantically interpreted as energy, and specifies (ii) in which direction the correlation holds. Now, suppose that we have a minimal musical sequence made of two musical events $M = \langle m_1, m_2 \rangle$, with m_1 being softer and m_2 being louder. We therefore have an increase in loudness in the music. All events which satisfy **the bolded rule**, i.e. all events that are such that they involve an increase in energy, will make that musical sequence true of them; and the set of all events which a musical sequence is true constitute the *denotation*¹⁴ of that musical sequence. In our case, the truth-condition can be stated as follows:

$M = \langle m_1, m_2 \rangle$ is true of $E = \langle e_1, e_2 \rangle$
if e_1 involves more energy than e_2 .

This is an extremely simplified version of a possible model of music semantics reduced to the interpretation of musical loudness, and several conditions have here been omitted for simplicity (for instance, we did not specify what the notion of energy was referring to exactly). Besides, many more musical properties are involved in the model developed by Schlenker (2018a) (in particular, truth conditions on time, pitch, and harmony), and even more should be added in the future. But this gives the idea: the purpose of music semantics is to identify musical properties that bear meaning and formalize the semantic rules that generate meaning from musical form (which we just did for loudness): in short, develop a model which can derive the truth conditions of any musical sequence.

One must note that although the ground concepts are shared between natural language semantics and music semantics, the two systems are very different with respect to at least two characteristics. First, music semantics is very abstract: a same music can be

¹⁴For now, we do not expand of what kind of things lie in the denotation; but we will be using the term ‘denoted’ applied to objects, events and situations as meaning ‘belonging to the denotation’

true of many different situations, and much more abstract than the semantics of natural languages. In our example, M is indeed true of any event gaining energy; that is arguably an extremely large set of possible denotations. Second, it is based on very different rules. In particular, music semantics heavily relies on (a particularly abstract form of) iconicity, i.e. the fact that a certain musical form will resemble its denotation.¹⁵

3.2 Musical inferences

We have just seen an example of how a given musical sequence with a determined arrangement of loudness levels of its sub-events can be true of certain events. As soon as we have a notion of musical truth, it is possible to define a related notion of inference. As a first approximation, we suppose that musical inferences are of the same sort as linguistic inferences. Let us go back to our previous example: we have seen that a musical event M with increasing loudness is true of an event E involving an increase in energy. Any event that makes M true will then also make a statement of the form ‘ E involves an increase in energy’ true. We will therefore say that M licenses the inference that the denoted event E involves an increase in energy.

Musical inferences are very diverse in nature: music can provide different kind of information about the extra-musical world. Schlenker (2017) argues that there are at least two categories of musical inferences. First, there are inferences which are lifted from normal auditory cognition and triggered by sound properties that are not exclusive to music: inferences from loudness, pitch, and timbre fall within that category. For instance, the fact that an increase in loudness licenses an inference about an object moving closer has nothing specifically musical to it: any sound with increasing loudness in the world would license an inference about the source of that sound coming closer. Second, there are inferences which are specific to music cognition, and which do not necessarily have counterparts in the non-musical world. Inferences triggered by harmony (i.e. the system of rules governing the hierarchical organization of chords within a pitch space structured around a centre) fall within that second category of inferences. It is possible that some

¹⁵Although there are cases of iconicity in language too, such as that reported in Guerrini (2020), they are quite rare in spoken languages, although much more common in Sign Languages Schlenker (2018a).

musical properties license inferences that fall in between these two categories. This may be the case of consonance. Although consonant musical events have been shown to be associated with more stability and pleasantness (Butler and Daston, 1968; Krumhansl et al., 2000), one could argue that consonance is found in non-musical sounds as well. In other words, the mechanisms that make listeners represent a given sound or aggregation of sounds as consonant might apply to both music and sounds; yet there might be music-specific ways to create consonance.

The results presented in this dissertation are essentially experimental results which investigate the empirical reality of both previous concepts. First, we will test whether listeners draw inferences about an extra-musical reality when listening to music, and we will provide experimental data which support this view. Second, we will be testing whether the semantic rules that define musical truth are accurate as stated, and we will see that empirical data calls for a few updates of the notion of musical truth and truth conditions. The last following section presents how we conducted this double inquiry.

4 Outline

The purpose of this dissertation is to provide insights as to how music semantics can be revealed through different experimental paradigms. In particular, it aims at investigating semantic effects from music in different environments. Because music semantics is very abstract, we reasoned that having music interact with other representational systems or other modalities should help reveal its semantic behavior in such contexts.¹⁶

It is first possible to analyze how the information conveyed by music behaves when interacting with a linguistic environment. In 4.1, we explain how we designed an experiment in which hybrid sentences containing music replacing words triggered inferences similar to that found in language alone. A second interaction we explored was that between music and visual animations, inspired by audiovisual tasks such as those in Blumstein et al. (2012), Sievers et al. (2013) or Eitan (2013). In 4.2, we present how we confirmed some semantic rules stated in Schlenker (2017) using audiovisual stimuli, and how this

¹⁶We are here using the intuition from Schlenker (2017) that “Semantic intuitions that would otherwise be very unclear can be sharpened by reducing the set of possible denotations.”

helped determine which kind of things music refers to. But one might in the end be curious to know what happens in music alone, i.e. when it does not interact with any other system. In 4.3, we introduce a paradigm that aimed at testing music semantics directly: by focusing on how music could evoke walking, we managed to establish some limitations of the current model of music semantics. Finally, we present in 4.4 a last experimental paradigm investigating how musical properties interact with body movement beyond formal interactions with language and vision. The present dissertation is structured in the very same way: to each of the following section corresponds a chapter which includes one or two research articles.

4.1 Meaning from music embedded in language

The first interaction we investigated was that of music with natural language. Taking inspiration from Tieu et al. (2019), who showed that gestures and visual animations replacing words in sentences could generate the same inferences as language alone, we built an experiment in which music were replacing words in sentences and showed that participants drew the same inferences as well. Consider the sentence in (1).

(1) a. Mary stopped smoking.

→ Mary used to smoke.

b. Did Mary stop smoking?

→ Mary used to smoke.

The meaning of sentence in (1)a. has two components: an at-issue component (i.e. the very topic of discussion), which is that Mary stopped to smoke, and a non-at-issue component, which is not directly conveyed and is taken for granted, namely that Mary used to smoke. Crucially, turning the sentence into a question cancels the at-issue component: one does not understand from (1)b. that Mary stopped smoking (it is precisely what is now being questioned), but it is still taken for granted that Mary used to smoke. This non-at-issue component, which resists question formation among other logical operations, is called a presupposition.

Tieu et al. (2019) showed that it was possible to trigger presuppositions by means of pro-speech gestures and pro-speech visual animations, i.e. gestures or visual animations that replace words in sentences. For instance, the question in (2), which involved the pro-speech gesture REMOVE_GLASSES standing for the gesture of the speaker removing their glasses, presupposes that student under question currently has glasses on.

(2) Will the student REMOVE_GLASSES?

→ The student currently has glasses on.

Our first contribution is to extend this paradigm to non-linguistic auditory stimuli, and to test whether the same mechanisms work with pro-speech music. In other words, we tested whether replacing words with music still triggers the same kinds of inference that are found in language alone, and that were shown to exist in hybrid sentences containing visual information. In the case of presuppositions, for instance, the paradigm we used was the following. Imagine a context in which some hikers are going up and down mountains, and one of them asks the question in (3).

(3) Do you think the next hiker will UPWARD_SCALE

→ The next hiker is at the bottom of a mountain.

Question (3) contains a musical stimulus UPWARD_SCALE, standing for an embedded ascending scale played by a harp and replacing the verb phrase in the question, and intended to be understood as meaning something like ‘go up’.¹⁷ It then seems that question (3) does trigger a presupposition about the location of the next hiker, which is (i) not directly conveyed by the question, and (ii) still taken for granted.

Just as in Tieu et al. (2019), our experiment investigated whether similar mechanisms could generate further inferential types besides presuppositions. Our results showed that it was possible to replicate this typology by pro-speech musical means. Together with Tieu et al. (2019), they support the idea that at least some semantic mechanisms accounting for linguistic inferences are not exclusive to language and that they can also integrate

¹⁷Crucially, and following Tieu et al. (2019), we made sure that the musical stimuli were not merely translated into words, for if it were, the paradigm would fail to demonstrate anything as to whether music is interpreted when embedded in sentences.

non-linguistic auditory information (be it musical or vocal) besides vision: processes underlying meaning creation in language are in fact not limited to linguistic items.

However, these results are not sufficient to establish the existence of a music semantics. Although they reveal that participants were able to retrieve some informational content from musical stimuli which they had never been exposed to before, these findings are consistent with other accounts that do not require semantic rules for the target inferences to be drawn. To get closer to the actual model of music semantics presented in section 3, we developed two further experimental paradigms, one involving an audiovisual task, and one involving a specific case-study. Both approaches aimed at restricting the investigation of musical meaning to cases where the semantics is more easily accessible to participants, either by forcing some semantic interpretations thanks to visual information, or by limiting the possible denotations to an identifiable event (in our case, that of someone walking).

4.2 Meaning from music interacting with visual scenes

The second paradigm we developed aimed at experimentally establishing some of the hypotheses provided in Schlenker (2017, 2022) about the interpretation of three musical properties, namely pitch, dynamics (or loudness), and timbre, through the interaction between musical sounds and visual animations. Although a great deal of work has been done on establishing correlations between these musical properties and several possible interpretations, these correlations have never been tested under explicit rules providing formal conditions on music.

By combining simple visual animations involving 3D objects and minimal musical stimuli, we created pairs of audiovisual scenes which were either compatible or incompatible with the predictions that followed from Schlenker (2017), and had participants select which of the two they preferred. We found that the results were broadly in line with the predictions of the model. In particular, participants systematically preferred the scenes that obeyed one of the following rules: (i) the higher the pitch, the higher the level of energy of the denoted object, and (ii) the louder the music, either the higher the level of

energy of the represented object, or the closer that object to a given viewpoint. A bit more surprisingly, we also found evidence in favor of the following rule: (iii) the higher the pitch, the closer the represented object. Although this association was not predicted and conflicts with previous research such as that from Eitan and Granot (2006), it is also in line with other experimental data (Eitan, 2013). We discuss further avenues for research in the corresponding paper. Importantly, these correlations were tested through pairings of dynamic changes in both musical properties and visual animations. This is essential because the theory predicts that *changes* in musical properties map onto *changes* in the denotation.¹⁸

A second part of our experiment was intended to demonstrate that not only were pitch and dynamics interpreted as energy and distance when paired with one object, but they can also serve to anchor reference to a specific object when multiple objects are present. Let us take an example to illustrate this. In the first part of the experiment, we established that listeners prefer to associate a crescendo (i.e. an increase in loudness) with an object moving closer or gaining energy, rather than moving away or losing energy (rule (ii)). Now, what happens if two objects are present in the visual scene? How do listeners integrate the auditory information with the visual information? By combining musical sounds with visual scenes involving two objects, we showed that listeners rely on the following rule: (iv) the musical sound refers to the object that satisfies the properties of that sound. In our example, only the object moving closer or gaining energy is associated with the sound. In our paper, we argue that this is an instance of *coreference* across media, in the sense that the denotation of the musical sound and that of the visual object both refer to a same object, though through different modalities (namely audition and vision).¹⁹ These results are consistent with predictions from theoretical work on reference relations in music (Schlenker, 2022), as well as with previous work on the existence of such relations in pictures (Abusch, 2013) and dance (Patel-Grosz et al., 2018).

A third and final part of this experiment aimed at clarifying the nature of the entities

¹⁸Although we showed in a later experiment, presented in 4.3, that this need to be systematic, this first approximation was needed to experimentally establish the reality of the semantic effects we tested.

¹⁹Note that there exists coreference within media as well, e.g. typically when two linguistic items such as two pronouns refer to a same individual.

music refers to. In section 3, we presented music semantics as relying on the core assumption that music can refer to the external world. In Schlenker (2017), music is claimed to convey information about non-musical objects; while in Schlenker (2022), this claim is updated to account for how music conveys information about situations. We thus tried to disentangle this point and provide an argument showing that musical properties are bound to properties of objects, and not just to properties of situations or scenes containing objects. To show this, we created timbre/object associations in audiovisual scenes and showed that the scenes in which that association was consistent were preferred. In short, it is not enough that some object in an audiovisual scene satisfies the properties of the musical sound: in our case, one and the same object has to satisfy both properties on timbre and dynamics to produce an acceptable audiovisual scene. Although subtle, this point is of particular relevance for subsequent developments in music semantics, as it suggests that each semantic rule should account for how a given musical property is interpreted at the situation level but also at the object level.

4.3 Musical meaning from music alone

Although an audiovisual paradigm was useful to establish some semantic associations from music by appealing to the visual modality, it is essentially unable to account for how music triggers inferences about possible objects when music is heard alone. The third paradigm we present here aimed at testing how listeners interpret music in the absence of any interaction with another system conveying meaning. But because of the abstract nature of music semantics (a given musical sequence can be true of many different objects), we still had to find a way to constrain the set of possible inferences that music licenses, and we developed a case-study on how music can represent someone walking. This case study was motivated by previous substantive work on (i) the correlations established between music and representations of movement, (ii) the intricate relationship between music and actual body movement, for instance in the case of dance, and (iii) the more specific historical and empirical relationship that holds between music and walking (such as in military marches).

This experiment had four main goals. First, it tested whether some music could evoke a walking character better than others. Under the assumption that music has an extra-musical meaning and can refer to objects through certain rules, our prediction was that it should. Based on introspective judgments and taking inspiration from musical sequences accompanying walking events (such as in cartoons), we built a reference stimulus which listeners rated as being more evocative than control stimuli which were hardly associated with the same event. Although insufficient to demonstrate that our stimulus directly encoded reference to a character walking, these preliminary findings showed that to the least, the stimulus licensed an interpretation as a walking character.

Second, the experiment aimed at identifying the musical formal properties responsible for this association. Through the rewriting of a short musical sequence and the minimal destruction of each identified property, we showed that at least five properties were involved in making reference to a walking character. For a musical sequence to optimally denote a walk, it needed to be reducible to a (i) steady (ii) alternation of (ii) two chords, which were both (iv) relatively minimally distant in the tonal space and (v) absolutely consonant. Participants systematically picked the stimulus satisfying all these properties as evoking a character walking the best, over any stimulus satisfying only a subset of these properties.

Third, we show that these properties could not all be derived from the model presented in Schlenker (2018*b*). As detailed in Section 3, this model is based on preservation rules which associate a change in a musical property to a change in the associated object property. For instance, an increase in loudness is associated with a decrease in distance to the object, or an increase in the energy level of that object. But these rules have been argued to be sometimes too weak to account for more fine-grained musical inferences (Migotti, 2019). In particular, this model does not make predictions as to whether some musical properties are interpreted in absolute terms, for instance whether listeners can directly map an atomic musical event with a given loudness level to a given distance or level of energy. Our experiment showed that there seem to exist such absolute mechanisms for the interpretation of consonance: it is for instance required that any physically stable

event is musically represented by a consonant musical event. Meeting the requirements on the relative consonance level of musical events representing the different steps of a walking event is not enough: each step, which is itself a physically stable event, has to be represented by a consonant musical event (such as consonant chord).

Fourth, this experiment proposed to experimentally test a last theoretical dimension of the theory, namely that each musical event has to be interpreted (as an event).²⁰ While in Schlenker (2017), the assumption was made that each musical event is interpreted as an event, we wanted to test our intuition that some musical events are semantically ‘empty’, and that they might act as modifiers on musical meaning instead of being fully interpreted as independent events. We thus enriched the stimulus used throughout the whole experiment by adding an extra-note in the musical sequence, resulting in a stimulus still evoking a walk but sounding bouncier, more light-hearted. Our prediction was that if it is true that some musical events are not interpreted, then the enriched stimulus would still license an inference about a character walking. Contrary to our predictions, we found that this minimal modification altered the semantic association with a walking character, suggesting that it changed musical meaning more significantly than predicted.

Together, these findings call for a more systematic investigation of case studies in which the semantic associations between some musical pieces and what they evoke in listeners are broken down into all musical properties and rules involved.

4.4 Music-induced body effects

Finally, once it has been established that it is possible to constrain listeners’ inferences by identifying rules mapping a musical sequence to a possible set of situations (such as situations in which someone is walking), one might finally be interested in whether the same sequences generate the same inferences in a listener who physically experiences walking. What happens when a listener starts walking on a musical sequence which also evokes walking? Some musical properties are indeed not only associated with properties of extra-musical objects, they also correlate with motor/physical effects. Some authors have

²⁰For a detailed discussion of the notion of event in music semantics and its theoretical implications, see Zaradkzi (2021).

claimed that musical meaning is even primarily grounded in bodily experience (Brower, 2000) or physiological experience (Antovic, 2009). On this view, characterizing some music as ‘tense’ simply results from an actual muscular tension. This line of work directly relates to investigations of unconscious physical reactions triggered by the recognition of acoustic patterns. For example, Bedoya et al. (2021) showed that applying an acoustic filter modeled after the modification that smiling induces in the voice to a violin sound not only makes that sound judged as happier, but also triggers subtle muscle reactions in listeners similar to those that occur when they actually smile. Others have called for the integration of a notion of ‘self-motion’ to explain how we generate motion representations from music: Clarke (2001) suggested that a possible reason for which listeners may feel movement when listening to music is because the processing of some pitch structures produces biological effects to the ear which are similar to those experienced when actually moving.

Although our first idea was to test how the stimuli used in the experiment presented in 4.3 shape listeners’ gait patterns, we realized that there was not enough data in the literature on the physical and motor reactions to changes in some musical properties. In particular, while a considerable amount of evidence had established music/motor time-related correlations (for instance, people adjusting their pace to match a musical beat, or more generally following changes in tempo), none was available as to the effect of time-unrelated dimensions (such as pitch height or harmony) on gait specifically. Komeilipoor et al. (2015) had however shown that dissonances were responsible for creating irregularities in a tapping task, leaving open the question of whether similar effects could be found for other movements. We therefore built a paradigm testing for the effect of pitch height on gait patterns using sequences of notes randomly varied in pitch on which participants were tasked to walk. We found that the lower the pitch, the longer the steps duration, and the stronger the stepping force applied onto the ground.

Note that these associations could occur in the absence of a music semantics: it does not seem to require any kind of semantic system for the physiological effect to arise in the first place. Still, the question of how these physical effects interact with musical

semantic representations is puzzling. A first possibility is that listeners simply try to mimic their semantic interpretation of musical sounds, as if music was understood to be about themselves. This could explain why they unconsciously ‘align’ their walking patterns with how they represent walking events corresponding to certain pitch ranges and patterns: listeners might map the pitch properties of a musical excerpt to properties of objects, and then somehow act as if those properties applied to themselves. For instance, one hears something that generates a representation of a big object, which then has motor effects akin to those which would occur if the listener themselves were bigger. We discuss some reasons for which such a far-fetched mechanism could exist in the corresponding paper, in particular in light of similar mechanisms pertaining to rhythmic synchronization (Styns et al., 2007). It is also possible that some musical properties (such as pitch) and some movements (such as walking) trigger mental representations which overlap: under this view, the semantics does not necessarily come first, and then affects the body, but both mechanisms could just interact at the same time. In any case, our findings are surprisingly consistent with some theories of pitch height stating that it encodes object size: we did find that participants stepped more heavily on low pitches, which have been shown to be associated with large objects (Eitan and Timmers, 2010). Although this paradigm only established a correlation from music to gait patterns, it raises interesting questions as to whether and how music semantics interacts with other music-induced phenomena, and opens fruitful research perspectives on how these effects can be used in gait rehabilitation.

5 Bibliography

- Abusch, D. (2013), ‘Applying Discourse Semantics and Pragmatics to Co-reference in Picture Sequences’, *Proceedings of Sinn und Bedeutung* **17**, 9–25.
- Antovic, M. (2009), ‘Towards the Semantics of Music: the 20th Century’.
- Bedoya, D., Arias, P., Rachman, L., Liuni, M., Canonne, C., Goupil, L. and Aucouturier, J.-J. (2021), ‘Even violins can cry: specifically vocal emotional behaviours also drive the perception of emotions in non-vocal music’, *Philosophical Transactions of the Royal Society B: Biological Sciences* **376**(1840), 20200396.
- Bernstein, L. (2005), *Leonard Bernstein’s young people’s concerts*, 1st amadeus press edn, Amadeus Press, Pompton Plains, N.J. OCLC: ocm61445826.
- Blacking, J. (1976), *How musical is man?*, Faber and Faber, London.
- Blumstein, D. T., Bryant, G. A. and Kaye, P. (2012), ‘The sound of arousal in music is context-dependent’, *Biology Letters* **8**(5), 744–747.
- Bowling, D. L., Gill, K., Choi, J. D., Prinz, J. and Purves, D. (2010), ‘Major and minor music compared to excited and subdued speech’, *The Journal of the Acoustical Society of America* **127**(1), 491–503.
- Bowling, D. L., Sundararajan, J., Han, S. and Purves, D. (2012), ‘Expression of Emotion in Eastern and Western Music Mirrors Vocalization’, *PLoS ONE* **7**(3), e31942.
- Brower, C. (2000), ‘A Cognitive Theory of Musical Meaning’, *Journal of Music Theory* **44**(2), 323–379. Publisher: [Duke University Press, Yale University Department of Music].
- Butler, J. W. and Daston, P. G. (1968), ‘Musical Consonance as Musical Preference: A Cross-Cultural Study’, *The Journal of General Psychology* **79**(1), 129–142.
- Clarke, E. (2001), ‘Meaning and the Specification of Motion in Music’, *Musicae Scientiae* **5**(2), 213–234.
- Cone, E. T. (1982), *The composer’s voice*, 1. paperback pr edn, Univ. of Calif. Pr, Berkeley.
- Cross, I. and Woodruff, G. E. (2009), Music as a communicative medium, in R. Botha and C. Knight, eds, ‘The Prehistory of Language’, Oxford University Press, pp. 77–98.
- Davies, S. (1994), *Musical Meaning and Expression*, Cornell University Press.
- Eitan, Z. (2013), How pitch and loudness shape musical space and motion, in ‘The psychology of music in multimedia’, Oxford University Press, New York, NY, US, pp. 165–191.
- Eitan, Z. and Granot, R. Y. (2006), ‘How Music Moves’, *Music Perception* **23**(3), 221–248.

- Eitan, Z. and Timmers, R. (2010), ‘Beethoven’s last piano sonata and those who follow crocodiles: Cross-domain mappings of auditory pitch in a musical context’, *Cognition* **114**, 405–422. Place: Netherlands Publisher: Elsevier Science.
- Godøy, R. I. and Leman, M., eds (2010), *Musical gestures: sound, movement, and meaning*, Routledge, New York. OCLC: ocn298781501.
- Granroth-Wilding, M. and Steedman, M. (2014), ‘A Robust Parser-Interpreter for Jazz Chord Sequences’, *Journal of New Music Research* **43**(4), 355–374.
- Guerrini, J. (2020), ‘Vowel quality and iconic lengthening’, *Proceedings of Sinn und Bedeutung* pp. 242–255 Pages. Artwork Size: 242-255 Pages Publisher: Proceedings of Sinn und Bedeutung.
- Hanslick, E. (1957), *The beautiful in music*, 8. print edn, The library of Liberal Arts, Indianapolis.
- Heider, F. and Simmel, M. (1944), ‘An Experimental Study of Apparent Behavior’, *The American Journal of Psychology* **57**(2), 243.
- Honing, H. (2003), ‘The Final Ritard: On Music, Motion, and Kinematic Models’, *Computer Music Journal* **27**(3), 66–72.
- Huron, D. (2006), *Sweet Anticipation: Music and the Psychology of Expectation*, The MIT Press.
- Ilie, G. and Thompson, W. F. (2006), ‘A Comparison of Acoustic Cues in Music and Speech for Three Dimensions of Affect’, *Music Perception* **23**, 319–329. Place: US Publisher: University of California Press.
- Juslin, P. N. and Laukka, P. (2003), ‘Communication of emotions in vocal expression and music performance: Different channels, same code?’, *Psychological Bulletin* **129**(5), 770–814.
- Juslin, P. N. and Sloboda, J. A., eds (2001), *Music and emotion: theory and research*, Series in affective science, Oxford University Press, Oxford ; New York.
- Katz, J. and Pesetsky, D. (2011), ‘The Identity Thesis for Language and Music’.
- Koelsch, S. (2012), *Brain and music*, Wiley-Blackwell, Chichester, West Sussex ; Hoboken, NJ. OCLC: ocn767563922.
- Komeilipoor, N., Rodger, M. W. M., Craig, C. M. and Cesari, P. (2015), ‘(Dis-)Harmony in movement: effects of musical dissonance on movement timing and form’, *Experimental Brain Research* **233**(5), 1585–1595.
- Krumhansl, C. L., Toivanen, P., Eerola, T., Toiviainen, P., Järvinen, T. and Louhivuori, J. (2000), ‘Cross-cultural music cognition: cognitive methodology applied to North Sami yoiks’, *Cognition* **76**(1), 13–58.
- Larson, S. (2012), *Musical forces: motion, metaphor, and meaning in music*, Musical meaning & interpretation, Indiana University Press, Bloomington. OCLC: ocn707212791.

- Lerdahl, F. and Jackendoff, R. (1983), *A generative theory of tonal music*, the MIT press, Cambridge, Mass. London.
- Maus, F. E. (1988), ‘Music as Drama’, *Music Theory Spectrum* **10**(1), 56–73.
- Mehr, S. A., Singh, M., Knox, D., Ketter, D. M., Pickens-Jones, D., Atwood, S., Lucas, C., Jacoby, N., Egner, A. A., Hopkins, E. J., Howard, R. M., Hartshorne, J. K., Jennings, M. V., Simson, J., Bainbridge, C. M., Pinker, S., O’Donnell, T. J., Krasnow, M. M. and Glowacki, L. (2019), ‘Universality and diversity in human song’, *Science* **366**(6468), eaax0868. Publisher: American Association for the Advancement of Science.
- Meyer, L. B. (1956), *Emotion and meaning in music*, paperback ed., [nachdr.] edn, Univ. of Chicago Press, Chicago. Ill.
- Migotti, L. (2019), Steps towards a theory of music semantics, Master’s thesis, Ecole Normale Supérieure.
- Monahan, S. (2013), ‘Action and Agency Revisited’, *Journal of Music Theory* **57**(2), 321–371. Publisher: [Duke University Press, Yale University Department of Music].
- Patel-Grosz, P., Grosz, P. G., Kelkar, T. and Jensenius, A. R. (2018), ‘Coreference and disjoint reference in the semantics of narrative dance’, *Proceedings of Sinn und Bedeutung* **22**(2), 199–216. Number: 2.
- Rohrmeier, M. (2011), ‘Towards a generative syntax of tonal harmony’, *Journal of Mathematics and Music* **5**(1), 35–53.
- Saslaw, J. (1996), ‘Forces, Containers, and Paths: The Role of Body-Derived Image Schemas in the Conceptualization of Music’, *Journal of Music Theory* **40**(2), 217–243. Publisher: [Duke University Press, Yale University Department of Music].
- Schlenker, P. (2017), ‘Outline of Music Semantics’, *Music Perception* **35**(1), 3–37.
- Schlenker, P. (2018a), ‘Iconic pragmatics’, *Natural Language & Linguistic Theory* **36**(3), 877–936.
- Schlenker, P. (2018b), ‘What is Super Semantics? *’, *Philosophical Perspectives* **32**(1), 365–453.
- Schlenker, P. (2022), ‘Musical meaning within Super Semantics’, *Linguistics and Philosophy* **45**(4), 795–872.
- Schulte-Tenckhoff, I. and Feld, S. (1988), ‘Sound and Sentiment. Birds, Weeping, Poetics, and Song in Kaluli Expression’, *Cahiers de musiques traditionnelles* **1**, 214.
- Sievers, B., Lee, C., Haslett, W. and Wheatley, T. (2019), ‘A multi-sensory code for emotional arousal’, *Proceedings of the Royal Society B: Biological Sciences* **286**(1906), 20190513.
- Sievers, B., Polansky, L., Casey, M. and Wheatley, T. (2013), ‘Music and movement share a dynamic structure that supports universal expressions of emotion’, *Proceedings of the National Academy of Sciences* **110**(1), 70–75.

- Slobin, M. (1993), *Subcultural sounds: micromusics of the West*, Music/culture, University Press of New England, Hanover, NH.
- Stravinsky, I. (1935), *An autobiography*, Norton, New York.
- Styns, F., Van Noorden, L., Moelants, D. and Leman, M. (2007), ‘Walking on music’, *Human Movement Science* **26**(5), 769–785.
- Tieu, L., Schlenker, P. and Chemla, E. (2019), ‘Linguistic inferences without words’, *Proceedings of the National Academy of Sciences* **116**(20), 9796–9801.
- Trevarthen, C. (1999), ‘Musicality and the intrinsic motive pulse: evidence from human psychobiology and infant communication’, *Musicae Scientiae* **3**(1-suppl), 155–215.
- Zaradzki, L. (2021), *Les événements en sémantique linguistique et musicale*, PhD dissertation, Université Paris Cité.

Chapter 2

Meaning from music embedded in language

Purpose

We start our experimental investigation of musical meaning by exploring how musical information behaves when music is embedded in language and used in lieu of words in sentences. Previous theoretical and experimental work had established that gestures and visual animations replacing words in sentences can give rise to different inferences from the linguistic typology. Here, we show that these findings extend to non-linguistic musical items: the mental mechanisms generating meaning from language can also integrate musical meanings when music is embedded in language.

The article contained in this chapter was published in the *Linguistics and Philosophy* special issue on Super Linguistics (2023). Reference is provided in the paper.

LINGUISTIC INFERENCES FROM PRO-SPEECH MUSIC

MUSICAL GESTURES GENERATE SCALAR IMPLICATURES, PRESUPPOSITIONS, SUPPLEMENTS, AND HOMOGENEITY INFERENCES *

Léo Migotti & Janek Guerrini
Institut Jean Nicod (ENS-EHESS-CNRS)

ABSTRACT

Language has a rich typology of inferential types. It was recently shown that subjects are able to divide the informational content of new visual stimuli among the various slots of the inferential typology: when gestures or visual animations are used in lieu of specific words in a sentence, they can trigger the very same inferential types as language alone (Tieu et al., 2019). How general are the relevant triggering algorithms? We show that they extend to the auditory modality and to music cognition. We tested whether pro-speech musical gestures, i.e. musical excerpts that replace words in sentences, can give rise to the same inferences. We show that it is possible to replicate the same typology of inferences using pro-speech music. Minimal and complex musical excerpts can behave just like language, gestures, and visual animations with respect to the logical behavior of their content when embedded in sentences. Specifically, we found that pro-speech music can generate scalar implicatures, presuppositions, supplements, and homogeneity inferences.

Keywords Semantics | Super semantics | Scalar implicatures | Presuppositions | Supplements | Homogeneity inferences | Iconicity

*This paper has been published: Migotti, L., Guerrini, J. (2023). Linguistic inferences from pro-speech music. *Linguistics and Philosophy*. <https://doi.org/10.1007/s10988-022-09376-9>

This research received funding from the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme (grant agreement No 788077, Orisem, 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-IDEX-0001-02 and FrontCog ANR-17-EURE-0017. This paper is part of the Special Issue "Super Linguistics", edited by Pritty Patel-Grosz, Emar Maier and Philippe Schlenker.

Contributions: JG designed the first stimuli, which LM generated and recorded. Subsequent stimuli were designed by LM and JG. JG and LM designed the experiment and wrote the paper. LM managed online data collection and statistical analyses.

Contents

1	Introduction	36
2	Methods	41
3	Replication of the typology of inferences	42
3.1	Scalar implicatures	42
3.1.1	In language and from visual stimuli	42
3.1.2	From music	43
3.1.3	Results	46
3.1.4	Discussion	46
3.2	Presuppositions	47
3.2.1	In language and from visual stimuli	47
3.2.2	From music	49
3.2.3	Results	50
3.2.4	Discussion	51
3.3	Supplements	54
3.3.1	In language and from visual stimuli	54
3.3.2	From music	55
3.3.3	Results	56
3.3.4	Discussion	57
3.4	Homogeneity inferences	58
3.4.1	In language and from visual stimuli	58
3.4.2	From music	59
3.4.3	Results	61
4	Iconicity controls	65
4.1	Experimental paradigm	65
4.2	Discussion	67
5	Conclusion	69
6	Acknowledgments	71
	Appendix I: Musical gestures and vocal gestures	76
	Appendix II: Statistical models	77

1 Introduction

Natural language semantics has long evidenced the existence of a rich typology of content, showing that the meaning of sentences can be split up into different content types (Schlenker, 2016; Kadmon, 2001). Let us take the example of presuppositions.² Sentence (1)a conveys two kinds of information: the speaker asserts that John does not smoke now, while taking for granted that John used to smoke. Such inferences, presuppositions, survive in different logical environments. When turning the sentence into a question in (1)b, the assertion that John stopped smoking is now questioned, but the presupposition that he used to smoke remains.

- (1) a. John stopped smoking.
 \rightsquigarrow John used to smoke.
- b. Did John stop smoking?
 \rightsquigarrow John used to smoke.

It has been argued that such a projection pattern (namely, here, the fact that the presupposition is preserved under question formation³) is lexically encoded, i.e. determined on the basis of word meaning (Levinson, 2000; Abusch, 2002). The meaning of “stop” would for instance be divided as in (2).

- (2) x stops Q -ing.
 - At-issue: x does not Q .
 - Presupposed: x Q -ed before.

Alternatively, some theorists have claimed that general algorithms can predict when an inference triggered by a given word is treated as a presupposition, in part because across languages the same projection patterns are observed (Simons et al., 2010; Tonhauser et al., 2013; Abusch, 2010; Abrusán, 2010; Schlenker, 2021). On this view, humans can divide asserted content from presupposed content productively and on the fly. Here, we propose to test the productivity of these mechanisms by making use of non-linguistic items such as music: if such general and productive procedures exist, then the same inferential patterns should arise

²Here we use the example of presuppositions as an illustration of the main argument of the paper. The argument however applies similarly to the three other inferences tested: scalar implicatures, supplements and homogeneity inferences. A summary of the different inferential mechanisms can be found in Appendix III.

³Generally, the projection *problem* refers to the computation of the presupposition and asserted content of a sentence from the presupposition and asserted content of its constituents. By contrast, projection *tests*, such as the family-of-sentences test (so-called in Chierchia and McConnell-Ginet (2000)), i.e. embedding under negation, modality or question, are used to check whether a given proposition is a presupposition (Heim, 1990, 1992; Stalnaker, 1974; Geurts, 1999; Beaver, 2001; Chemla, 2009)

from non-linguistic stimuli. In other words, we infer the existence of a hardwired, general-purpose algorithm that subdivides meaning into at-issue and non-at-issue from people’s ability to endorse inferences containing both types of content on the fly. This can be conceived of as a variant of the argument of the poverty of the stimulus (PoS) (Chomsky, 1980, 1988). PoS states that the sentences children are exposed to when learning a language do not contain enough of the information needed to develop a thorough understanding of the grammar of the language. Hence, learning is at least partly made possible by a core of innate linguistic mechanisms.

Similarly for our case: contextual conditions such as knowledge of specific words of a given language cannot be responsible for systematic inferential behavior in presence of stimuli subjects have plausibly never been exposed to. Consequently, there is no real alternative to the conclusion that humans have an algorithm available to adequately arrange content in at-issue and non-at-issue before experience with a given linguistic item. In both cases, systematic behavior (competent use of a language for the general PoS, systematic inferential behavior in our case) is not backed by enough exposure to justify its learning, and thus can only be caused by subject-internal structures (Universal Grammar for the general PoS, general triggering algorithms in our case).

Both theoretical and experimental research on pro-speech gestures and pro-speech visual animations (i.e. gestures or visual animations replacing words in sentences) have corroborated these predictions. Schlenker (2019a) developed a typology of embedded gestural depictions replacing, following, or co-occurring with words, showing that, in particular, the content of pro-speech gestures can be divided among familiar slots of the inferential typology. Tieu *et al.* showed this experimentally. Consider (3) from Tieu *et al.* (2019):

- (3) a. The student REMOVE_GLASSES.
 ↪ The student currently has glasses on.
- b. Will the student REMOVE_GLASSES?
 ↪ The student currently has glasses on.

Here, REMOVE_GLASSES stands for the gesture mimicking someone removing their glasses. Sentence (3)a presupposes that, at the time of the utterance, the student has glasses on. Similarly, the question in (3)b still presupposes that the student has glasses on, just like the presupposition that John used to smoke was preserved under question in (1). Subjects were shown to be able to access this presuppositional content significantly more than control inferences. In general, Tieu *et al.* found that embedded visual stimuli can give rise to the same inferential types as purely linguistic ones. This suggests that the algorithm humans apply to divide at-issue content from non-at-issue content must be productive, and, because the stimuli were visual, it cannot be limited to language narrowly intended.

How general is the algorithm that divides at-issue content from non-at-issue content in the case of presuppositions, and how general are other inferential algorithms producing, for instance, scalar implicatures, supplements, or homogeneity inferences? Here we ask whether they extend to the auditory modality and to music cognition. We used pro-speech music, i.e. musical excerpts that replace words in sentences, henceforth *musical gestures*.⁴ Building our paradigm after Tieu et al. (2019), we tested whether musical gestures can give rise to the same inferential typology as language, pro-speech gestures and pro-speech visual animations. For example, in a context involving a person that was hiking, sentence (4) behaved just like a change-of-state verb like “stop”, and triggered the presupposition that the hiker is down a mountain both when embedded in a declarative sentence as in (4)a, and when it is embedded in an interrogative sentence as in (4)b.⁵

- (4) a. The hiker will [UPWARD_SCALE](#).⁶
 ~> The hiker is at the foot of a mountain.
- b. Will the hiker [UPWARD_SCALE](#)?
 ~> The hiker is at the foot of a mountain.

⁴Although ‘pro-speech music’ [literally, music replacing words] is a specific kind of musical gestures [i.e. the iconic musical motives or excerpts roughly used in lieu of gestures in Tieu et al. (2019)], we use ‘pro-speech music’ and ‘musical gestures’ interchangeably throughout this paper.

⁵Except for the paradigm testing supplements in section 3.3, we mainly used basic scales, drum sounds or isolated tones. Our definition of these as music could be contested because of their simple nature. However, even if these stimuli did not count as music, our claims on the generality of the algorithm that divides content in at-issue and non-at-issue would remain unaltered.

⁶All musical gestures can be directly accessed by clicking on the hyperlinks.

UPWARD_SCALE stands for a minimal musical stimulus of a classical C major scale played by a harp. Each note of the scale was played one after the other following a standard rise in frequency at a 990 notes/s rate to approach the sound of a real harp *glissando*, often used in classical music. In general, because of the intuitive mapping between gesture and meaning (Schlenker, 2019a), the informational content of iconic gestures can be grasped even in absence of previous exposure (Schlenker, 2017). The reader has plausibly never been exposed to (4)a or (4)b, but the content of the musical gesture UPWARD_SCALE can still be productively divided between at-issue content, *viz.* the hiker is going up, and presuppositional content, *viz.* at the time of utterance, the hiker is located at a low point in space.⁷ Crucially, the inference goes through even though (4)b is a question, which is a classical test for presupposition projection. Besides presuppositions, we try to replicate the typology of inferences already replicated with gestures and visual animations in Tieu et al. (2019), including scalar implicatures, supplements, and homogeneity inferences.

Here we are merely interested in the semantic and pragmatic behavior of the informational content of musical sounds. This paper does not aim at deciding between different theories of the respective inferences we tested; rather, it reports experimental data that any theory should account for.⁸ Still, in some cases these results make it possible to exclude a significant class of theories, as for instance in the case, mentioned above, of lexical theories of presuppositions.

Neither is the paper about how the content of pro-speech music relates to the actual musical properties of the stimuli, i.e. about musical meaning and how it is derived. Still, we provide first insights into how this meaning can interact with the logical structures of language, which suggests the existence of a non-trivial informational content in music. We leave open the challenging question of whether music triggers more sophisticated content, especially outside of a linguistic context.

⁷As pointed out by an anonymous reviewer, we cannot know when precisely participants drew the inference. However, whenever the inference is actually triggered, it is unclear how our predictions would be different at this stage. As the same issue was present in Tieu et al. (2019), we assume that our predictions would not have been different.

⁸For instance, our data does not allow us to decide between a grammatical or a neo-Gricean theory of scalar implicatures, but allows us to claim that regardless of the details of the algorithm responsible for scalar implicatures, this algorithm extends to music cognition and must therefore be domain-general.

Before we move on, let us address a potential worry concerning the methodology used here and in Tieu et al. (2019). There is the possibility that pro-speech music is systematically translated into words. In this case, there would be the possibility that the relevant inferences arise because they are lexically encoded in the words of the translation. Consequently, our results would be uninformative about the generality of the algorithm dividing content in at-issue and non-at-issue. Two reasons militate against this hypothesis. First, just as in Tieu et al. (2019), subjects of our experiment were able to interpret fine-grained gestural iconic information that was absent from the words of the closest verbal translation. Second, such verbal translations would make iconic dimensions at-issue when not encoded in a verbal translation that lexically makes them non-at-issue, a behavior that seems to be excluded by the logical tests we provide in section 4.

This paper is structured as follows. In section 2, we detail the experimental setup. In section 3.1 we present the paradigm we used to test for scalar implicatures. Two stimuli involving respectively one and three repetitions of a drum sound competed and gave rise to scalar implicatures. In section 3.2, we present the paradigm we used to test for presupposition projection, already introduced in (4). While the presupposition was preserved under question formation, participants did not behave as expected in the classical test under “none”. We discuss in detail possible explanations of the difference between our results and Tieu *et al.*’s. We then move on to other linguistic inferences and introduce a paradigm testing for supplements in section 3.3. Pro-speech music did indeed behave like a supplement, yielding the typical conditional projection. In section 3.4, we test homogeneity inferences from pro-speech music. There are different theoretical accounts of such inferences: some attribute the existence of homogeneity inferences to the noun phrase, others to the predicate. We prove “by case” that pro-speech music gives rise to homogeneity inferences both when it replaces the noun phrase and when it replaces the predicate. Finally, in section 4, we discuss in detail why it is unlikely that pro-speech music is systematically translated into words.

2 Methods

Stimuli were recordings of French sentences with musical gestures (either specifically generated for our purposes or pre-existing) either artificially generated through GarageBand (version 10.3.5) or taken from real music replacing one or some words.

We collected data on 68 participants. We excluded non-native French speakers, participants who had already taken one of our pilot experiment and participants who had failed the two attention checks, which left us with 53 subjects. Participants were recruited through the French platform Crowdpanel, and online informed consent was obtained for each of them. After a short training on three examples of hybrid sentences containing music, and for each stimulus, each participant had to assess to what extent the presented inference followed from the auditory stimulus (i.e. the spoken sentences with embedded music), using a slider bar ranging from 0, labeled *totally disagree* (in French: *pas du tout d'accord*), to 100, labeled *totally agree* (in French: *tout à fait d'accord*) (cf. Figure 1)⁹. Stimuli played automatically for each trial, but participants could listen to the stimulus as many times as needed. The experiment was set up on Qualtrics. All stimuli were fully randomized. In each section, we report comparisons of generalized mixed-effects models of the data using R (version 3.6.2.) to assess the contribution of each factor of interest to the model (R core Team (2016); Barr et al. (2013)). The details, justifications, simplifying procedures and coefficients of the models are available in Appendix II. The details of the stimuli are given in the next sections, for each type of inference.¹⁰

⁹As pointed out by a reviewer, it cannot be ruled out that when asked about inferences, subjects understand that they must guess the right word in the stimuli, e.g. ‘climb’. Indeed, the target sentence was visible while listening to the stimulus, just like in (Tieu et al., 2019). Since this issue applies to all of this literature, we leave it as a problem for future research to understand if and how participants consider the lexical material from the target sentence to be relevant to the comprehension of the stimulus.

¹⁰All material including stimuli, design files and analysis scripts are accessible at https://osf.io/hw45u/?view_only=89f983db777f49e9a6f5b41b3dea60d6. Material was uploaded prior to the beginning of the data collection - See PREREGISTRATION folder. Final results and statistics are available in the RESULTS folder.

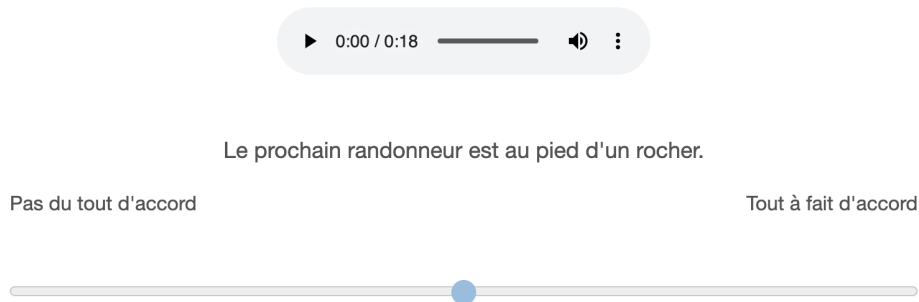


Figure 1: An example of a stimulus discussed in 3.2

3 Replication of the typology of inferences

3.1 Scalar implicatures

3.1.1 In language and from visual stimuli

Scalar implicatures convey an implicit, strengthened meaning beyond the explicit or literal meaning of an utterance. Originally, they were viewed as arising from pragmatic reasoning on the speaker’s communicative intentions (Grice, 1975). According to neo-Gricean approaches (Horn, 1972; Sauerland, 2004; van Rooij and Schulz, 2004; Chierchia et al., 2012), if a sentence S evokes S' and S' is more informative than S (or not less informative than S , depending on the specific approach), speakers infer that S' is false. For instance, in (5)a the speaker chose to utter “some” instead of the logically stronger “all”, and is therefore inferred to mean *some but not all* (Sauerland, 2012).

More recently, some theorists have argued that the mechanism responsible for scalar implicatures is instead a grammatical one (Chierchia et al., 2012). On this view, a silent operator with a meaning very similar to *only* is applied to scalar sentences: the inference in (5)c is the result of interpreting “some” in (5)a as *only some*, as shown in (5)b.

- (5) a. He read some of the books.
 b. *Exh* (He read some of the books.)
 \rightsquigarrow c. He read some, but not all of the books.

Either theory has at its core competition among alternatives interacting with informativity considerations.¹¹ Scalar implicatures were recently shown to arise between alternatives realized by means of a gesture or a visual animation provided in a context (Tieu et al., 2019). For instance, in the positive environment in (6) where TURN_WHEEL stands for the gesture mimicking a driver turning a wheel, we understand that TURN_WHEEL does not only mean *to turn* but *to turn somewhat, but not a lot*. TURN_WHEEL competes with the more informative gesture TURN_WHEEL_COMPLETELY, which is not used, thus taken to be false.

(6) a. He will TURN_WHEEL_COMPLETELY.
 ~> He will turn the wheel completely.

b. He will TURN_WHEEL.
 ~> He will turn the wheel, but not completely.

The inference in (6)b. could be explained by the fact that TURN_WHEEL semantically means *to turn somewhat, but not completely*. In this case, however, we would expect *not*-TURN_WHEEL to mean *to not turn somewhat*.

(7) a. He will not TURN_WHEEL_COMPLETELY.
 ~> He will turn the wheel, but not completely.

b. He will not TURN_WHEEL.
 ~> He will not turn the wheel at all.

We rather understand from (7)b that the wheel was not turned at all. The informativity pattern gets reversed under negation, just like in language. *not*-TURN_WHEEL is now logically stronger than *not*-TURN_WHEEL_COMPLETELY. The inferences in (7) show that there is competition between the two gestures.

3.1.2 From music

We reasoned that realizations with different numbers of repetitions of a sound could be represented by speakers as a logical scale (Horn, 1972). We generated two different realizations

¹¹For presentational clarity, we choose to present scalar implicatures by going through the Gricean reasoning. We then mention the alternative theory, the grammatical account of scalar implicatures. In any case, these should be viewed as placeholders for any theory of scalar implicatures. Note that any theory of scalar implicatures predicts that if alternatives are provided in the context, implicatures should be derived. This is, in a way, a sanity check to confirm that the inferential mechanisms work as expected with pro-speech sounds, and that there can be competition among musical excerpts.

of the timpani sound DRUM (from GarageBand instruments library): the weaker DRUM×1 and the stronger DRUM×3. We first set up a context introducing the stronger alternative:

- (8) **Context (positive environment):** Jean boxes regularly at the gym. During last week’s workout, he had a lot of energy, and was able to DRUM×3.

In the target premise (9), DRUM×1 evokes the more informative alternative DRUM×3, introduced in the context in (8):

- (9) **Target premise**

This week, Jean will DRUM×1.

Original stimulus

Cette semaine, Jean va DRUM×1.

This week, Jean will DRUM×1.

If DRUM×1 forms a scale with DRUM×3, the stronger alternative DRUM×3 should be taken to be false if not realized. We expect speakers to draw the inference that John will box somewhat, but not a lot:

- (10) **Target inference**

This week, Jean is going to box somewhat, but not a lot.

In the control stimulus in (11), we did not vary the number of repetitions:

- (11) **Control premise**

This week, Jean will DRUM×3

Original stimulus

Cette semaine, Jean va DRUM×3.

This week, Jean will DRUM×3.

Because in (11) the informativity of the stimulus is not manipulated, we expect speakers to stick with the context and endorse the baseline inference in (12) when given the control premise.

- (12) **Baseline inference**

This week, Jean will box a lot.

Just like for linguistic stimuli, the critical test is negation. Without such a test, the expected inference from the positive environment in (12) may be explained otherwise. Namely, DRUM×3 may convey that Jean will box a lot simply because it means *to punch something exactly three times*, and not because it constitutes a logically stronger alternative than

DRUM×1. To rule out this alternative explanation, we set up a context in which both alternatives were introduced:

- (13) **Context (negative environment):** Jeanne is boxing at the gym. At last week’s session, she had a lot of energy, and was able to DRUM×3. But during the second week of training, she did not DRUM×1.

If the two realizations form a scale, under negation, informativity should be reversed: *not*-DRUM×1 constitutes a more informative alternative with respect to *not*-DRUM×3:

- (14) **Target premise**
This week, Jeanne will not DRUM×3.

Original stimulus

Cette semaine, Jeanne ne va pas DRUM×3.
This week, Jeanne NEG will not DRUM×3.

If DRUM×3 means *to punch something exactly three times*, we expect its negation to mean *not punch something exactly three times*, i.e. two or less, or four or more. If on the other hand the two realizations form a scale, as we argue, we expect *not*-DRUM×3 to convey that although not a lot, some boxing still occurred, as in (15) below.

- (15) **Target inference**
This week, Jeanne will box somewhat, but not a lot.

And similarly for the weaker realization: if DRUM×1 means *to punch something exactly once*, then *not*-DRUM×1 should mean *not punching something exactly once*. This yields an inference pattern different from what is expected under our hypothesis that *not*-DRUM×3 forms a scale with *not*-DRUM×1. In this case, *not*-DRUM×1 should convey that there was no boxing at all.

- (16) **Control premise**
This week, Jeanne will not DRUM×1.

Original stimulus

Cette semaine, Jeanne ne va pas DRUM×1.
This week, Jeanne NEG will not DRUM×1.

We thus expect *not*-DRUM×1 to be interpreted as *did not box at all* rather than *did not punch something exactly once*, which would allow for boxing more than once:

- (17) **Baseline inference**
This week, Jeanne will not box at all.

3.1.3 Results

A scalar implicature triggered by the target premise would lead participants to endorse the target inference more than the baseline inference. Such a difference could be due to an *a priori* preference for the target inference, independently of the target premise. For that reason, we looked at the interaction between target *vs* control premise and target *vs* baseline inference. We did find a significant interaction between the two factors that rule out this possibility in both environments ($\chi^2 = 170$, $p < 0.001$ in the positive environment; $\chi^2 = 21$, $p < 0.001$ in the negative environment), compatible with the triggering of a scalar implicature resulting from the competition of two musical alternatives, namely DRUM×1 and DRUM×3.

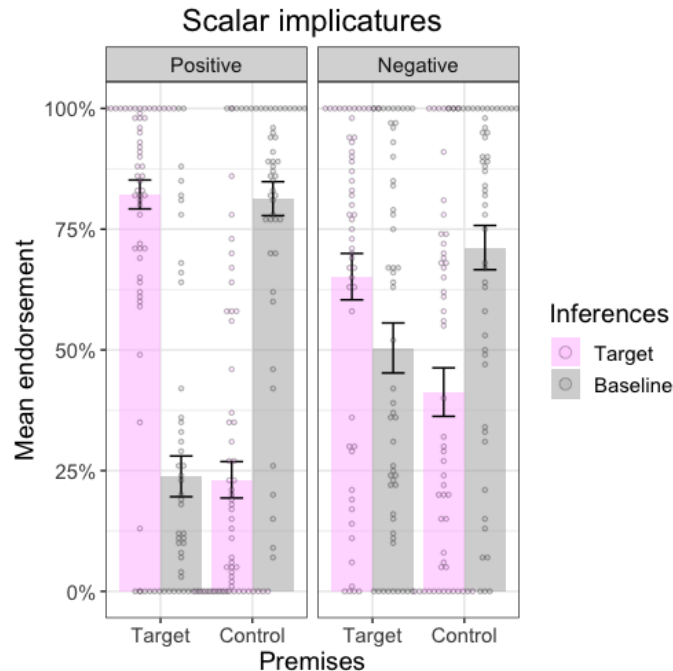


Figure 2: Mean endorsement rate for scalar implicatures

3.1.4 Discussion

In the positive environment, as expected, DRUM×1 competes with the more informative alternative DRUM×3 so that DRUM×1 is understood as *to box somewhat but not a lot*. In the negative environment, *not-DRUM×3* competes with *not-DRUM×1*. Thus *not-DRUM×3* was interpreted as *box somewhat, but not a lot*. This shows that when two realizations have different logical informativity, they compete with each other just like alternatives compete in

the case of gestures and visual animations in Tieu et al. (2019), leading to the generation of scalar implicatures by means of non-purely linguistic means.¹²

3.2 Presuppositions

3.2.1 In language and from visual stimuli

A presupposition is a meaning component that is taken for granted that must follow from the local context of an expression, and which sentences inherit across different logical operators, such as negation (Stalnaker, 1974; Heim and Kratzer, 1998). In language, some particular expressions are associated with the triggering of presuppositions. For instance, in (18), the presupposition that John used to smoke is triggered by “stop”. In general, change-of-state verbs (e.g. “start”, “become”, “stop”) all presuppose their initial state (Abrusán, 2010). For John to be able to stop smoking, John must have been smoking to begin with.

Presuppositional content is untouched by logical operations that change the meaning of a sentence, such as negation, modality, and questions (Abusch, 2010; Abrusán, 2011; Chemla, 2009). For instance, the question in (18)b still presupposes that John used to smoke.

- (18) a. John stopped smoking.
 \rightsquigarrow John used to smoke.
- b. Did John stop smoking?
 \rightsquigarrow John used to smoke.

Tieu et al. (2019) found that presuppositions could also be generated by means of pro-speech gestures and visual animations. First, they were preserved under question formation:

- (19) **Context** Aliens are green. But when they are in a meditative state, their antennae are blue. There is a meditation session in progress on the first floor of an architecture firm. Jane is watching the union representatives and says:

¹²In our stimuli, alternatives were provided in the context: the alternative, e.g. DRUM×3, to the musical gesture, DRUM×1, was salient in the context. For this reason, our results don’t speak to the issue of alternative generation. Our point is that scalar implicatures can be triggered by pro-speech music as long as two alternatives are available. In other words, whatever the mechanism responsible for alternative generation, subjects are able to interpret one musical alternative as logically stronger than the other, have it compete with the other, and finally draw a scalar implicature from this process. However, if scalar implicatures were in fact to arise when alternatives are not directly given, then a general theory of alternatives could be needed. This is an exciting question for future research. Schlenker (2020) discussed such a theory in relation to Katzir (2007) in the case of gestures. As to music, it would not be surprising that implicatures are triggered whenever a musical gesture competes with a contextually salient (but not explicitly given) more informative alternative.

(20) “Will the union representatives’ antennae GREEN_TO_BLUE?”¹³
(animation content: **bar is green at first, then slowly whole bar goes blue**)

↪ The union representative is currently not in a meditative state.

→ The union representative is currently in a meditative state.

The context used in (19) introduces the two states between which the transition denoted by the visual animation GREEN_TO_BLUE can be made. When turning from green to blue, the embedded visual animation triggers the presupposition that the initial state denoted by the green color holds before the change of state takes place, i.e. the individual is not in a meditative state at the moment of utterance.

Evidence for presuppositions being triggered by visual animations was found under “none”, too, as shown in (21) below.

(21) “None of the union representatives’ antennae will GREEN_TO_BLUE”

↪ None of the union representatives are currently in a meditative state.

→ Some of the union representatives are currently in a meditative state.

Tieu *et al.*’s (2019) choice to test for presupposition under “none” instead of under negation was motivated by Chemla (2009) results on strong universal inferences under “none” in French. This choice has significant theoretical advantages. Some apparent presuppositions that project out of negative environments such as (22)a. may be due to a scalar implicature and may not be strong evidence for presuppositional content. In (22) TURN_WHEEL_COMPLETELY stands for the gesture mimicking a driver turning a wheel completely. The fact that the agent is in front of a steering wheel may come as a consequence of the scalar implicature that although there was no complete turning of the wheel, there was still some turning (Tieu *et al.*, 2019). For there to be some turning, the agent must be in front of a wheel in the first place.

(22) He did not TURN_WHEEL_COMPLETELY.

a. The agent is in front of a steering wheel.

Under "none", however, scalar implicatures project existentially, not universally, as shown in (23) below (Chemla, 2009).

¹³The original sequence of sentences and the visual animations can be accessed at Tieu *et al.*’s supplementary materials page: <https://mfr.au-1.osf.io/render?url=https://osf.io/v5xa3/?direct%26mode=render%26action=download%26mode=render>

- (23) No student read all the books.
↗ (At least) one student read (at least) some of the books..

The universal projection under ‘none’ found by Tieu et al. (2019) and reported in (21) thus constituted very strong evidence in favor of presupposition projection.

3.2.2 From music

We tested whether pro-speech music could generate presuppositions. Consider the context below:

- (24) *Context*: Some hikers are hiking in the mountains, where there can be significant drops and peaks in elevation. They alternate between reaching the top and the foot of steep rocks. Two of them, who finished first, are talking while waiting for the others.

We used an upward scale played by a harp to evoke a hiker going up a mountain, as in (25).

- (25) **Target premise (question)**
One asks the other: “Will the next hiker `UPWARD_SCALE`?”

Original stimulus:

L’ un demande à l’ autre : “Est ce que le prochain randonneur va
The one asks to the other : “Is it that the next hiker will
`UPWARD_SCALE`?”
`UPWARD_SCALE`?”

Question formation is a classical test for presupposition (Chemla, 2009; Tieu et al., 2019). In (25), if the initial state of the musical scale is presupposed rather than at-issue, the sentence should presuppose that the hiker is at the bottom of a rock. We thus expect the target inference in (26) to be more endorsed than its negation, the baseline inference in (27).

- (26) **Target inference**
The next hiker is at the foot of a rock.
- (27) **Baseline inference**
The next hiker is not at the foot of a rock.

Tieu *et al.* (2019) chose to test universal projection under “none”, and we follow suit to have a clear point of comparison between acoustic and visual gestures.

(28) **Target premise under “None”**

One tells the other: “None of the hikers will UPWARD_SCALE”.

Original stimulus

*L’ un dit à l’ autre : “Aucun des randonneurs ne va UPWARD_SCALE.
The one tells to the other : “None the hikers NEG will UPWARD_SCALE.*

Given Tieu *et al.*’s 2019 results, we thus expected the universal inference in (29) to be more endorsed than its negation, the baseline inference in (30).

(29) **Target inference**

Each of the hikers is at the foot of a rock.

(30) **Baseline inference.**

Not all of the hikers are at the foot of a rock.

3.2.3 Results

As explained in Appendix II, here we expected a significant effect of inference type on the endorsement of the premise: the inference containing the presupposition was expected to be significantly more endorsed than the baseline inference negating the presupposition. Under question formation, as expected, we found a significant effect of inference type ($\chi^2 = 6$, $p < 0.05$), compatible with the triggering of a presupposition. We also found a significant effect of inference type in the ‘none’ environment, ($\chi^2 = 22$, $p < 0.001$), where, a bit more surprisingly, the baseline inference that did not contain the presupposition was more highly endorsed than its counterpart containing the projected presupposition. We discuss some hypotheses regarding this pattern in the following subsection.

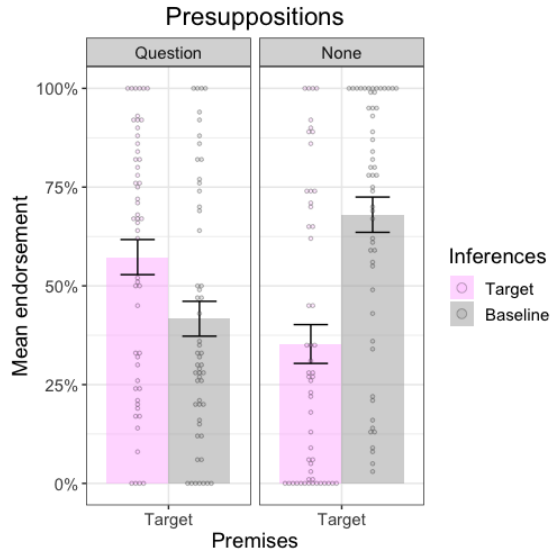


Figure 3: Mean endorsement rate for presuppositions

3.2.4 Discussion

In the interrogative environment, as expected, participants judged the initial state of UPWARD_SCALE to be true at the time of the utterance. Participants computed the initial state as non-at-issue, a behavior that matches the inferential pattern of change-of-state verbs. To wit, if the initial state had been at-issue, the question would have targeted the information that the hiker is at the bottom, amounting to something like “Is the hiker at the bottom and will he go up?”. Instead, the question only targeted the change of state, asking something like “Will the hiker go up?”. This supports our hypothesis that there is a productive algorithm at work in these cases, as subjects had never been exposed to such a sentence.

By contrast, UPWARD_SCALE embedded under “none” did not give rise to the expected universal inference that “Each of the hikers was at the foot of a rock”. We discuss several reasons for this observed pattern below.

First, it is worth mentioning that the data on projection under “none” are in part disputed: initially, Heim (1988) predicted universal projection, while Beaver (1994) predicted existential projection. Recent approaches find a complex picture (Zehr et al., 2016). We briefly review the theoretical insights and empirical evidence on this topic from Chemla (2009). Consider sentence (31), but see Zehr et al. (2016) for a complete review.

- (31) None of the players stopped smoking.
- a. At least one player used to smoke and none of them stopped smoking.
 - b. All of the players used to smoke and none of them stopped smoking.
 - c. None of the players both used to smoke and stopped smoking.
 - d. None of the players (who smoke) stopped smoking. \models All of the players who smoke smoke. \models At least one player smokes and none of them stopped smoking.

Although all theories of presupposition can derive more or less straightforwardly both types of projection, there is disagreement on whether the existential or the universal is the by-default reading. Sentence (31) could project weakly, as in ((31) a.), as predicted by a class of theories that stress existential projection (Beaver, 1994; Van Der Sandt, 1992; Geurts, 1998; Mandelkern, 2016).

Other theories predict that sentences like (31) display a default universal presupposition projection (Heim, 1988; Schlenker, 2008, 2010; George, 2008; Fox, 2013; Chemla and Schlenker, 2012; Mayr and Sauerland, 2016). The presence of existential projections is accommodated by these theories, as any existential can be captured as the result of a (conditional) domain restriction on the “none” - “none of the players (who smoked) stopped smoking”, as in (31)d., where the universal, adequately restricted, gives rise to an inference equivalent to an existential.

A third possible reading is a consequence of local accommodation, as in (31)c. Presuppositions can be interpreted within the scope of a logical operator that makes them at issue. In (32)a., the presupposition is interpreted within the scope of negation and the resulting truth conditions end up amounting to the conjunction of the supposed presuppositional content with the supposed at-issue content as in (32)b.

- (32) a. John didn’t stop smoking, he didn’t even start!
 b. John didn’t begin and stop a smoking habit, he didn’t even start!

The empirical evidence is not more uniform. As mentioned above, Chemla (2009) showed the availability of a universal reading in *none* sentences. In similar studies, it was argued that the most robust inference is the existential one (Sudo et al., 2012; Geurts and van Tiel, 2016). All studies found endorsement for both kinds of inferences, and Zehr et al. (2016) proposed

that both the existential and the universal inference are accessible to subjects, leaving open the question of the typology of the logical behaviors of presuppositions under *none*.

Turning to our result, it seems that a positive result would have been strong evidence of our hypothesis (all universal projections are presuppositions). However, in view of the complexity of the behavior of presupposition triggers under “none”, while the lack of universal projection weakens a bit our conclusion, it does not (at all) refute it (not all non-universal projections are non-presuppositions/not all presuppositions project universally).

A second point of discussion is that in spite of all this, the question remains of why we found a result different from Tieu et al. (2019). We here give two tentative explanations.

First, consider again our context in (24), reported below in (33):

- (33) Some hikers are hiking in the mountains, where there can be significant drops and peaks in elevation. They alternate between reaching the top and the foot of steep rocks. Two of them, who finished first, are talking while waiting for the others.

It is possible that participants inferred that not all of the hikers were at the bottom of a rock because the two characters discussing were at the top of a rock already, given that the context states that *they have already finished*. Let us again look at sentence (28), reported below in (34).

- (34) **Target premise (question)** One tells the other: “None of the hikers will **UPWARD_SCALE?**”

If the domain of “the hikers” in the target inference in (29) *includes* the two hikers discussing, the target inference in (29) should be false. In Tieu et al. (2019), instead, it was clear that the speaker and the object matter were distinct.

One further possible explanation for the observed difference is that it might be easier to convey absolute initial states in the visual modality than in the auditory modality. In Tieu et al. (2019), the visual animation of a green bar becoming blue (**GREEN_TO_BLUE**)

was used to refer to the color change in the antennae of aliens introduced in the context. In our own paradigm, the first note of the musical scale was used to refer to the level at which the hiker was located in space. While in our context any point on the rock can be denoted by the first note of the scale, the initial state “green” in Tieu et al.’s stimulus can only be interpreted as absolute, i.e. no other point on the color spectrum is compatible with this initial state. Then, some of the hikers may fail to climb up while (i) being in a position to climb up, i.e. not at the top, and (ii) not being at the foot of the rock (but higher). This would explain why subjects thought that not all of the hikers were at the foot of a rock.

3.3 Supplements

3.3.1 In language and from visual stimuli

Supplements are inferences triggered by non-restrictive relative clauses that behave like independent sentences (Potts, 2004; Schlenker, 2019a). For instance, sentence (35)a below is understood as (35)b and not as (35)c.

- (35) a. It is unlikely that Robin lifts weights, which is harmful.
 b. It is unlikely that Robin lifts weights. Lifting weights is harmful.
 c. It is unlikely that Robin lifts weights and that this is harmful.

Tieu et al. (2019) found that both gestures and visual animations used in lieu of a non-restrictive relative clause gave rise to the conditional presupposition characteristic of supplements. For instance, in (36), the gesture HIT replaces a non-restrictive relative clause similar to “which involves hitting”. The informational content of the gesture is not at-issue here, i.e. it is not targeted by “if”, hence the inference that if the event under question happens, it will involve some hitting.

- (36) If June bugs a classmate today - HIT, she will get a detention.
 \rightsquigarrow If June bugs a classmate today, it will involve hitting her.

By contrast, in (37), the content of the gesture is made at-issue by the use of “and does so *like this*”. The way in which the bugging happens, i.e. by hitting, is now made at-issue and thus targeted by the conditional, hence the weaker inference that hitting will not necessarily be involved.

- (37) If June bugs a classmate today and does so like this - HIT, she will get a detention.
↪ If June bugs a classmate today, it will not necessarily involve hitting her.

3.3.2 From music

We used an excerpt of real film music by American composer Bernard Herrmann from the soundtrack of the movie *Psycho* to denote a scary dog coming closer to the character in the context shown in (38).¹⁴ We reasoned that a rich musical gesture with a non-ambiguous emotional content could be informative about the scene while being interpreted as a non-restrictive relative clause.

- (38) **Context:** Marie is walking back home. She spots a dog on the other side of the street. She sometimes worries about dogs, because some of them can be vicious - **PSYCHO**.

In the target premise in (39), **PSYCHO** behaves like a non-restrictive relative clause. It conveys that if the dog comes closer, it will look somewhat dangerous or scary.

- (39) **Target premise**
If the dog comes to her - **PSYCHO**, Marie will cross the street.

Original stimulus

Si le chien s'approche d'elle - PSYCHO, Marie changera de trottoir.
If the dog self approaches to her - **PSYCHO**, Marie changes of sidewalk.

If the musical gesture **PSYCHO** logically behaves like a non-restrictive relative clause, we expect the target sentence in (39) to give rise to the target inference in (40), in which the behavior of the dog is not at-issue, thus not targeted by the conditional.

- (40) **Target inference**
If the dog comes to Mary, it will look vicious.

To ensure that the musical gesture **PSYCHO** was not interpreted as a non-restrictive relative clause in the control premise in (41), we used the deictic “this” in “and does so like **this - PSYCHO**” to make the informational content of the musical gesture at-issue.

¹⁴In this paradigm, we embedded complex music in language and uncovered rich linguistic inferences. A non-trivial extension of our paradigm may in future test if such logical inferences can arise in purely musical environments.

(41) **Control premise**

If the dog comes to her and does so like this - PSYCHO, Marie will cross the street.

Original stimulus

Si le chien s'approche d'elle et qu'il le fait comme ça -
If the dog self approaches to her and that it-NOM. it-ACC. does like this -
PSYCHO, *Marie changera de trottoir.*
PSYCHO, Marie changes of sidewalk.

The use of “this” now makes the informational content of the musical gesture at-issue and thus targeted by the conditional “if”, leading to the inference in (42) that the dog will not necessarily look vicious. Just like Tieu et al. (2019), we opted for a simpler formulation of the weaker control inference in (42) instead of the more convoluted exact negation of the target inference, “It is not the case that if the dog comes to Mary, it will look threatening.”

(42) **Baseline inference**

If the dog comes to Mary, it won't necessarily look vicious.

3.3.3 Results

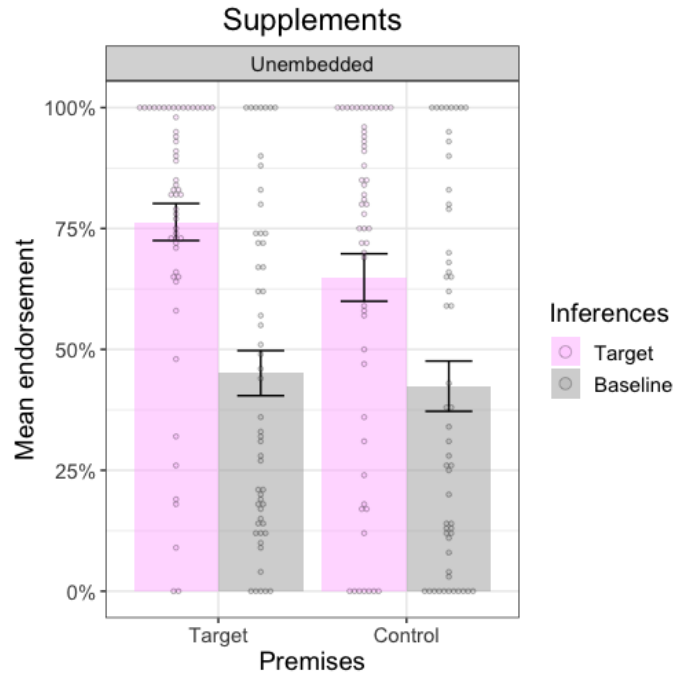


Figure 4: Mean endorsement rate for supplements

As explained in Appendix II, we are here interested in the interaction between inference type (supplement *vs* no supplement) and premise type (pro-speech gesture *vs* ‘like this’ control): to ensure that the stronger endorsement of the target inference containing the supplemental

information we observe for the target premise was not due to a default preference for this kind of inference regardless of the premise, we contrasted this premise with a control in which the musical gesture was made at-issue through the use of ‘like this’. Despite the higher endorsement of the target inference in response to the target premise, the interaction between the two factors was not significant ($\chi^2 = 1.5, p = 0.2$). In Tieu et al. (2019), only a marginal interaction was found for supplements with visual animations. This lack of interaction might be due to the simplification of the negated inference used as a baseline, as we used “won’t necessarily...” instead of the full negation “It is not the case that, if the dog comes to Mary, it will look vicious”, which truth-conditions might well be different. We discuss a further difference with the original paradigm from Tieu et al. (2019) in the following subsection.

3.3.4 Discussion

The paradigm we used was slightly different from the test for supplements in Tieu et al. (2019), as we chose to introduce the content of the music explicitly in the context, while in Tieu et al. (2019), the gesture HIT mimicking someone boxing only occurred in the target sentences. We reasoned that introducing the content of the gesture would not affect its logical behavior. On the one hand, indeed, introducing the content of a pro-speech gesture would have jeopardized the validity of the results for inferences that can be easily lexically encoded, such as presuppositions under certain theories. On the other hand, because non-restrictive relative clauses are sentence-level constructions, there is not even the possibility that they are encoded in the meaning of a specific word.

Our paradigm did not aim at showing that subjects could derive a meaning from music but, rather, that they could treat a musical gesture as a non-restrictive relative clause when given its meaning. We thus introduced this meaning explicitly in the context, so that participants did not reject the target inference only because they perceived the excerpt as conveying a feeling other than fear. In Tieu et al.’s, the gesture HIT mimicking someone hitting someone else was used in lieu of a non-restrictive relative clause. It is quite unambiguous that HIT

means something like *hit*. On the other hand, the meaning of the musical gesture, an excerpt from real complex film music, would have been hardly the same for all participants.¹⁵ ¹⁶

3.4 Homogeneity inferences

3.4.1 In language and from visual stimuli

Homogeneity inferences arise from definite plural noun phrases behaving universally in positive sentences, but existentially under negation (Križ, 2015; Križ and Spector, 2020; Spector, 2013; Löbner, 2000; Križ, 2016; Gajewski, 2005). Sentence (43)a has a universal reading: we understand that Mary found all of her presents. Under negation, however, the definite plural has an existential reading (i.e. “at least one”). Sentence (43)b is understood as *Mary did not find any of her presents*, and not as *Mary found some, but not all of her presents*, which involves the logical negation of “all of her present”.

- (43) a. Mary found her presents.
 ↪ Mary found all of her presents.
- b. Mary did not find her presents.
 ↪ Mary did not find any of her presents.

Tieu et al. (2019) investigated whether pro-speech gestures and visual animations used in lieu of both a verb *and* a definite plural noun phrase could generate homogeneity inferences. For instance, in (44) and (45), the gesture TAKE-2-HANDED-RIGHT replaced both a verb meaning *take* and a definite plural. The definite component was realized by means of pointing, and the plural component was realized by means of the iteration of a gesture representing a coin, which is a common way of signaling plural in sign languages (Pfau and Steinbach, 2006).

¹⁵Although this music unambiguously conveys a feeling of fear, danger or suspense, there are many possible situations this music can refer to that would trigger such feelings, considering that music can indeed refer to several external non-musical situations sharing some structural and/or emotional properties (Schlenker, 2017, 2019b)

¹⁶As music is not generally used to directly communicate ideas or convey information about the world, investigating the intermediary case of onomatopoeias may bridge our findings with the results from Tieu et al. (2019). Since onomatopoeias are used in combination with language and are highly iconic, there are substantial reasons to believe that any inference type from the typology we replicated with musical gestures could just as well be replicated with onomatopoeias: since onomatopoeias occur naturally in lieu of words, we would expect them to display the same inferential behavior as musical gestures which do *not* occur naturally in lieu of words and are arguably more difficult to process. To verify that this was the case, we ran a similar experiment with onomatopoeias instead of musical stimuli. Results were not significant for all inference types (see Appendix I for details). However, the results were not significantly different across the two experiments, leaving open the possibility that the onomatopoeias experiment, which we ran on fewer participants, was lacking power.

In the positive environment in (44), the pro-speech gesture triggers the universal inference that all coins were indeed taken.

- (44) Sam will TAKE-2-HANDED-RIGHT.
 \rightsquigarrow Sam will take all of the coins.

Under negation, sentence (45) gives rise to an existential reading of the gesture TAKE-2-HANDED-RIGHT (“at least one coin”) instead of the universal reading found in the positive environment (“all of the coins”). *Not*-TAKE-2-HANDED-RIGHT is interpreted as *not at least one coin*, i.e. no coin at all, rather than *not all of the coins*, i.e. some but not all of the coins.

- (45) Sam will not TAKE-2-HANDED-RIGHT.
 \rightsquigarrow Sam will not take any coin.

While most research on homogeneity inferences focuses on the role of the noun phrase, more recent work has suggested that these inferences rather originate in the predicate associated with the noun phrase (Križ, 2019). We thus proceeded by case: in section 3.4.2, we report the results for stimuli in which the musical gesture replaced a noun phrase, while in section 3.4.3 we report a parallel paradigm but with the musical gesture in lieu of the predicate.

3.4.2 From music

Pro-speech music in lieu of the noun phrase While the two components of the definite plural were visually realized in Tieu et al. (2019), it was not possible to realize them both in the auditory modality. The auditory counterpart of iteration, marking plural, was straightforward, and we used the iteration of a same musical note to mark plural. Crucially, however, we could not find an auditory counterpart of pointing to reproduce the definite component. We thus chose to focus on manipulating the iterative component only and on verifying that the musical gesture was interpreted as a definite plural.

In (46), HARP \times 3 stands for three repetitions of a same harp sound evoking three harp players in an orchestra, while FLUTES \times 3 stands for three repetitions of a same flute sound evoking three flute players in the same context.

- (46) **NP Positive Context:** Every Thursday, the students of a music school are gathering, and the conductor chooses some of them to play for the evening concert. Tonight, three harp-players - HARP×3 and three flute-players FLUTE×3 are present.

We performed a first test in a positive environment, as shown in (47).

- (47) **Target positive premise**
The conductor made HARP×3 play.

Original stimulus

Le chef d' orchestre a fait jouer HARP×3.
The chief of orchestra has made play HARP×3.

The musical gesture HARP×3 is expected to have a universal reading and be interpreted as *all harps*, leading to the inference in (48), which we expect to be significantly more endorsed than its negation, the baseline inference in (49).

- (48) **Target inference**
All harps played.

- (49) **Baseline inference**
Some but not all of the harps played.

Negation constitutes a crucial test to correctly explain the pattern observed in the positive environment, for two reasons. First, homogeneous expressions behave as existentials in negative environments. Second, an alternative explanation of the pattern we observe for the positive environment is that HARP×3 elicits a numeral, i.e. *exactly three harps*. Under negation, the two explanations come apart.

If HARP×3 is understood as a numeral, i.e. *three harps*, its negation *not-HARP×3* should mean *not exactly three harps*, i.e. two or less, or four or more. This is compatible with the control in (52). If, by contrast, HARP×3 is understood as a definite plural, we expect it to be interpreted existentially under negation, i.e. *not-HARP×3* is interpreted as *not at least one harp*, i.e. no harp at all.

- (50) **Target negative premise**
The conductor did not make HARP×3 play.

Original stimulus

Le chef d' orchestre n' a pas fait jouer HARP×3.
The chief of orchestra NEG has not made play HARP×3.

If HARP×3 is interpreted as a definite plural, then we expect a higher endorsement of the homogeneous target sentence in (51) resulting from an existential reading of the musical gesture under negation, compared to the baseline inference in (52) resulting from a universal reading of the musical gesture (i.e., “The conductor did not make *all harps* play”).

(51) **Target inference**
No harp played.

(52) **Baseline inference**
Some but not all of the harps played.

3.4.3 Results

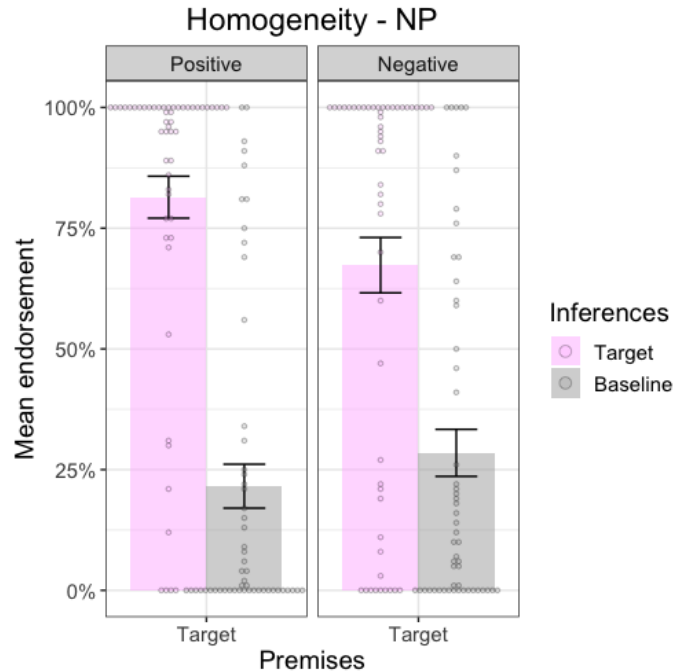


Figure 5: Mean endorsement rate for homogeneity (NP)

As explained in Appendix II, if homogeneity can be created by musical gestures standing for definite plurals, we expected an effect of inference type (homogeneous target *vs* non-homogeneous baseline): we expected the target homogeneous inference to be significantly more endorsed than its negation in both the positive and negative environments. This is what we find: the homogeneous inference (‘all harps played’) was preferred in response to the positive premise ($\chi^2 = 96, p < 0.001$), and similarly, the homogeneous inference (‘no harp played’) was preferred in response to the negative premise ($\chi^2 = 24, p < 0.001$). This

suggests that it is possible to generate homogeneity inferences with musical gestures in lieu of definite plurals.¹⁷

Pro-speech music in lieu of the predicate Recent work suggested that homogeneity inferences might not be systematically linked to a definite plural noun phrase, but rather to the kind of predicate associated with the noun phrase (Križ, 2019). If definite noun phrases were responsible for homogeneity inferences, it would be difficult to explain how non-homogeneous readings arise from sentences like (53).

- (53) The students are numerous.
a. \rightsquigarrow ?? Each of the students is numerous.

Križ (2019) argues that predicates, rather than bare plurals, are responsible for the presence or absence of homogeneity. While in Tieu et al. (2019), the test for homogeneity involved the gesture TAKE_ALL encompassing both the predicate TAKE and a definite plural noun phrase THE_COINS, the test we presented in section 3.4.2 only involved a musical gesture replacing the noun phrase.

Since there is the theoretical possibility that homogeneity comes from the predicate, we also tested whether sentences in which pro-speech music is inserted in lieu of the *predicate* can give rise to homogeneity inferences as well. In (54), the musical gesture MARSEILLAISE featuring a children’s choir singing the French national anthem was used to evoke the action of singing.

- (54) **Positive Context (predicate)**
The students of a class are learning to sing the Marseillaise - MARSEILLAISE.

We applied the same protocol as in section 3.4.2. We performed a first test in a positive environment, where we expected sentence (55) to give rise to the universal inference in (56).

¹⁷As pointed out by the Editor, instead of considering whether the premise triggered a homogeneous or a non-homogeneous reading, we could have shown that the interpretation of the musical gesture flips from universal in the positive environment to existential in the negative environment. In this case, we would have been interested in the interaction between the inference type factor, and the environment (positive *vs* negative). We display this alternative analysis in Appendix II, and show that it leads to a significant interaction between the two factors, indicating that the reading did change from universal to existential once embedded under negation.

(55) **Target positive premise**

On the 14th of July, the class has MARSEILLAISE.

Original stimulus

Le jour du 14 juillet, la classe a MARSEILLAISE.

The day of 14th July, the class has MARSEILLAISE.

(56) **Target inference**

The children in the class all sang the Marseillaise.

Just as in 3.4.2, the target inference in (56) needs to be contrasted with the non-universal baseline inference in (57).

(57) **Baseline inference**

Some but not all of the children in the class sang the Marseillaise.

We then performed a second test in a negative environment, i.e. the predicate MARSEILLAISE was in the scope of negation, as in (58).

(58) **Target negative premise**

On the 14th of July, the class did not MARSEILLAISE.

Original stimulus

Le jour du 14 juillet, la classe n' a pas MARSEILLAISE.

The day of 14th July, the class NEG has not MARSEILLAISE.

The negation of the musical predicate (*not*-MARSEILLAISE) now applies homogeneously to each member of the set denoted by the group singular noun phrase “the class”, leading to the inference in (59) that no child in the class sang the Marseillaise, which we expected to be highly endorsed.

(59) **Target inference**

None of the children in the class sang the Marseillaise.

By contrast, we expected a low endorsement of the logical negation of this inference, i.e. the weaker baseline inference in (60) that some but not all of the children in the class sang the Marseillaise.

(60) **Baseline inference**

Some but not all of the children in the class sang the Marseillaise.

In this paradigm, none of the tested inferences involve a definite plural noun phrase, so that the effect of the musical predicate could be isolated if there is one. If the sentence gives

rise to a homogeneity inference, we expect a strong endorsement of the target inferences (i.e. a universal reading of the sentence in the positive environment, and an existential one under negation) compared to the baseline inference.

An intuitive objection is that these inferences are not due to the predicate but to the group singular noun phrase “the class”. In this case, however, the homogeneity inference originates in the noun phrase and thus our paradigm presented in 3.4.2 provides the relevant evidence. If, on the other hand, homogeneity originates in the predicate, the results presented in this section provide evidence that pro-speech music predicates can give rise to homogeneity inferences.

The results show a significant effect of inference type in both the positive ($\chi^2 = 96, p < 0.001$) and the negative environments ($\chi^2 = 24, p < 0.001$). This suggests that it is possible to give rise to homogeneity inferences by means of a musical gesture replacing the predicate instead of the definite plural noun phrase.

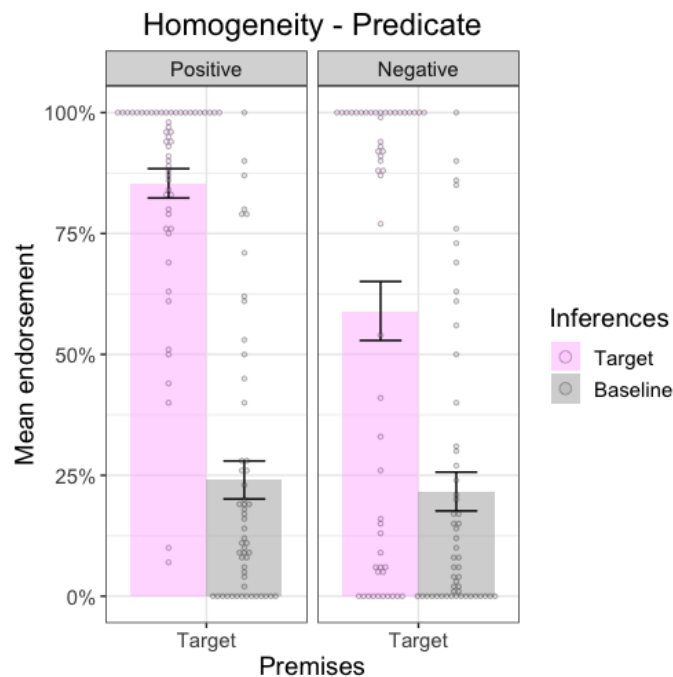


Figure 6: Mean endorsement rate for homogeneity (predicate)

4 Iconicity controls

4.1 Experimental paradigm

As mentioned in section 1, an alternative explanation of our data is that musical gestures gave rise to the observed inferences simply because they are systematically translated into words. Note that Tieu *et al.* faced the very same concern with pro-speech gestures and visual animations, which was also addressed by including iconicity controls throughout the experiment.

To rule out this alternative explanation, we included 8 iconicity trials randomly displayed throughout the experiment. We manipulated these stimuli to check whether participants were able to interpret their modification iconically. To this end, we used two different realizations of an upward scale, a normal and a fast one, to refer to a hiker climbing either slowly or quickly. We used two other realizations, a very long musical scale and a shorter one, to evoke the height of the mountain, in the same context as the one used in 3.2. The target sentences were then paired up with a matching or a mismatching inference. If participants were able to retrieve the expected iconic informational content from the musical gesture, we expected a significantly higher endorsement of the matching inference for the target premise with respect to the mismatching inference.

For the first iconicity control in (61), `LONG_UPWARD_SCALE` stands for a C major scale played over a wide pitch range of four octaves (i.e., four full upward scales, from middle pitch to very high), used to evoke a very high mountain.

(61) Target premise

The next hiker will `LONG_UPWARD_SCALE`.

Original stimulus

Le prochain randonneur va `LONG_UPWARD_SCALE`.

The next hiker will `LONG_UPWARD_SCALE`.

We expected participants to be able to (i) perceive the difference between the two realizations of the upward scale and (ii) interpret them iconically in terms of the height of

the mountain in the context. Consequently, we expected a high endorsement of the inference containing the matching iconic component in (62).

(62) **Matching inference**

The mountain the hiker is about to climb is high.

In (63), UPWARD_SCALE stands for the same scale played at the same speed but over a narrower pitch range of two octaves (i.e., two full upward scales only, from middle pitch to moderately high), to evoke a mountain of average height. Note that this musical gesture was the one used in 3.2 to test for presupposition projection. Our prediction was that the longer the scale, the higher the inferred height of the mountain.

(63) **Control premise**

The next hiker will UPWARD_SCALE.

Original stimulus

Le prochain randonneur va UPWARD_SCALE.

The next hiker will UPWARD_SCALE.

We expected that this neutral realization of the scale would convey information about the mountain being of average height, and thus a high endorsement of the inference in (64), in which the iconic component allowing subjects to assess the mountain to be very high was absent.

(64) **Mismatching inference**

The mountain the hiker is about to climb is not very high.

For the second iconicity control, we contrasted the baseline realization of the scale UPWARD_SCALE in (67) evoking a neutrally fast hike with SLOW_UPWARD_SCALE in (65), in which the pitch range remained the same but the speed was slowed down four times to evoke a slower climbing of the mountain.

(65) **Target premise**

The next hiker will SLOW_UPWARD_SCALE.

Original stimulus

Le prochain randonneur va SLOW_UPWARD_SCALE.

The next hiker will SLOW_UPWARD_SCALE.

Similarly, participants should be able to both perceive and interpret the difference between the two realizations iconically in terms of the hiking speed, and we expected a high endorsement of the inference in (66) containing the iconic component.

(66) **Matching inference**

The hiker will climb the mountain slowly

(67) **Control premise**

The next hiker will UPWARD_SCALE.

Original stimulus

Le prochain randonneur va UPWARD_SCALE.

The next hiker will UPWARD_SCALE.

The iconic component allowing for the inference about a high hiking speed was absent from this control premise, which we expected to give rise to the inference in (68).

(68) **Mismatching inference**

The hiker will climb the mountain fast.

We found a significant effect of inference type ($\chi^2 = 63$, $p < 0.001$ for the first iconicity control, $\chi^2 = 28$, $p < 0.001$ for the second). This means that listeners were able to both perceive subtle musical changes and interpret them iconically.

4.2 Discussion

We have shown that musical gestures convey fine-grained iconic information. Specifically, the pitch range of the stimulus was interpreted as the height of the mountain, and the tempo at which the upward scale was played was interpreted as the hikers' speed. It is very unlikely that these musical gestures were merely translated in words, because all these iconic dimensions are absent from reasonably simple translations. For instance, the information about height conveyed by UPWARD_SCALE is absent from a simple translation as "go up". It could still be, however, that more complex translations are involved. These would have to be extraordinarily rich: "The hiker will LONG_UPWARD_SCALE" would have to be translated as something like "The hiker will climb a mountain of such-and-such height at such-and-such a speed...". However, even if pro-speech music were indeed translated into such long chains of words, the music-as-translation theory makes wrong predictions on the projection patterns

of these iconic dimensions. Consider a translation of the target premise which, for ease, we simplify as (69).

(69) The hiker will climb up the mountain slowly.

The music-as-translation theory predicts that all pieces of fine-grained information iconically conveyed be *systematically at-issue* when not *lexically* encoded as non-at-issue. Any musical gesture conveying fine-grained iconic information should therefore systematically give rise to a scalar implicature. That is, SLOW_UPWARD_SCALE would mean *climb slowly*, thus its negation *not-SLOW_UPWARD_SCALE* should give rise to the scalar implicature that there is still some climbing involved, given that the more informative *not-UPWARD_SCALE* was not uttered. This prediction seems right if we look at the negation of (69), as in (70).

(70) The hiker will not climb up the mountain slowly.

↪ The hiker will not climb up the mountain slowly, but will still climb up the mountain.

The premise in (70) does not, however, trigger the same inference when using the musical gesture LONG_UPWARD_SCALE.

(71) The hiker will not LONG_UPWARD_SCALE.

a. ↪ The hiker will not climb up the mountain slowly, but will still climb up the mountain.

b. ↪ The hiker will not climb the mountain at all, but if she did, she would have done so slowly.

A first possible reading of the sentence in (71) is the same scalar implicature found in (71)a. However, another possible reading is (71)b, where the iconic information paraphrasable as “slowly” is not at-issue anymore. While this reading is not the most obvious, it becomes completely salient in context. Consider the following example:

(72) **Context:** Mary and John are playing a snail race, and they want their snails to climb all sorts of obstacles.

Mary tells John: “Your snail will not SLOW_UPWARD_SCALE.”

↪ John’s snail will not climb the next obstacle but if it had climbed it, it would have done so slowly.

→ John’s snail will not climb the next obstacle slowly, but will still climb it.

This inference is a typical cosupposition (Schlenker, 2018a) of the form *x did not Q, but if x did Q, x would have Qed like this*. Crucially, it is not triggered in the purely linguistic example in (70), as shown in (73).

- (73) The hiker will not climb up the mountain slowly.
→ The hiker will not climb the mountain at all, but if she did, she would have done so slowly.

If the musical gesture is only translated into words and no other non-translatable iconic enrichment is involved, then there is at least one available reading which cannot be explained. Consequently, there is no real alternative to the conclusion that there is a non-verbal component to pro-speech music. It is worth noting that similar inferences were found to be triggered by co-speech sounds and auditory animations (Tieu et al., 2018), which make non-at-issue contributions to the sentence.

5 Conclusion

The purpose of this study was to replicate the inferential typology tested on pro-speech gestures and visual animations of inferences in Tieu et al. (2019) with pro-speech music. Participants behaved systematically in response to novel stimuli, suggesting that a general cognitive algorithm is responsible for the appropriate arrangement of content, rather than a purely linguistic procedure relying on word meaning. Specifically, our results show that pro-speech music can give rise to four types of inferences.

We showed that musical gestures can assume different levels of informativity and compete, resulting in the computation of scalar implicatures (cf: section 3.1).

Moreover, we showed that pro-speech music can convey presuppositions. Presupposed content projected under question (cf. section 3.2), but did not project universally under “none”. While this difference with Tieu et al. (2019) weakens our hypothesis a bit, it does not constitute negative evidence given the complex behavior of presuppositions under “none”. Testing simple negation will provide a helpful test to show whether the observed behavior under “none” is specific to “none” or due to general complex projection patterns of pro-speech music.

Conditional inferences typical of supplements, i.e. non-restrictive relative clauses, can also arise from rich pro-speech music (cf. section 3.3). However, although the target inference with the supplemental information was more highly endorsed than the control inference without this information, we did not find the expected interaction between premise type and inference type. As Tieu et al. (2019) reported a marginal interaction in the case of pro-speech visual animations, further research will be needed to understand whether abstractness might weaken the triggering of supplements.

In section 3.4, we showed that pro-speech music can give rise to homogeneity inferences, both when replacing a definite plural noun phrase and when inserted in lieu of a predicate. Section 4 finally illustrated why it is very unlikely that these results can be explained by the mere translation of musical gestures into words.

Together, these findings suggest that the mechanisms responsible for these inferences and for the division of meaning along different slots of the inferential typology are (i) productive and (ii) general purpose. Indeed, (i) participants were able to draw the expected inferences in absence of previous exposure with the musical gestures, and (ii) they were able to operate logical computations on auditory, non-linguistic stimuli.

Finally, the question of the existence of a music semantics, recently raised in Schlenker (2017, 2019b) (a.o.), is a very exciting and promising one, but our results do not allow us to make any claim about musical meaning. However, we think that the evidence we provide about the interaction between language structure and musical content constitutes a first step to better understand what musical meaning is and how it behaves logically. In particular, the extent to which some of the observed inferences may arise in music alone, i.e. not embedded in language, is an inspiring question for future research.

6 Acknowledgments

We are greatly indebted to Philippe Schlenker and Emmanuel Chemla for in-depth discussion of virtually every aspect of this project. We also thank Amir Anvari for his helpful comments on the theoretical underpinnings of this work and suggestions for improvement on the first draft. Thank you to Salvador Mascarenhas, Rob Pasternak and Lyn Tieu for discussion of the early versions of this work. Finally, we would like to thank the audiences of the ‘Linguistic investigations beyond language’ workshop at ZAS, GLOW 2019, and the Linguae seminar.

Compliance with Ethical Standards

The authors have no financial or non-financial interest to disclose. Data collection was approved under Opinion number 20-733 of the Institutional Review Board of the French Institute of medical research and Health (IRB00003888, IORG0003254, FWA00005831). Online informed consent was obtained for all participants prior to the beginning of the experiment.

References

- Abbott, B. (2000). Presuppositions as nonassertions. *Journal of Pragmatics*, 32(10):1419–1437.
- Abrusán, M. (2010). Triggering verbal presuppositions. 20:684–701.
- Abrusán, M. (2011). Predicting the presuppositions of soft triggers. *Linguistics and Philosophy*, 34(6):491–535.
- Abusch, D. (2002). Lexical Alternatives as a Source of Pragmatic Presuppositions. *Semantics and Linguistic Theory*, 12:1.
- Abusch, D. (2010). Presupposition Triggering from Alternatives. *Journal of Semantics*, 27(1):37–80.
- Barr, D. J., Levy, R., Scheepers, C., and Tily, H. J. (2013). Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language*, 68(3):255–278.
- Beaver, D. (1994). When variables don’t vary enough. In *Semantics and linguistic theory*, volume 4, pages 35–60.
- Beaver, D. I. (2001). Presupposition and Assertion in Dynamic Semantics. *Stanford: CSLI Publications*.
- Breheny, R. (2019). Scalar Implicatures. In Cummins, C. and Katsos, N., editors, *The Oxford Handbook of Experimental Semantics and Pragmatics*, pages 38–61. Oxford University Press.
- Chemla, E. (2009). Presuppositions of quantified sentences: experimental data. *Natural Language Semantics*, 17(4):299–340.
- Chemla, E. and Schlenker, P. (2012). Incremental vs. symmetric accounts of presupposition projection: an experimental approach. *Natural Language Semantics*, 20(2):177–226.
- Chierchia, G. (2006). Broaden Your Views: Implicatures of Domain Widening and the “Logicity” of Language. *Linguistic Inquiry*, 37(4):535–590.
- Chierchia, G., Fox, D., and Spector, B. (2012). Scalar implicature as a grammatical phenomenon. In *Semantics: An International Handbook of Natural Language Meaning*. De Gruyter Mouton, Berlin, claudia maienborn, klaus von heusinger, and paul portner edition.
- Chierchia, G. and McConnell-Ginet, S. (2000). *Meaning and grammar: An introduction to semantics*. MIT press.
- Chomsky, N. (1980). On Cognitive Structures and Their Development: A Reply to Piaget. In *Language and Learning*. Routledge and Kegan Paul, London, piatelli-palmarini, m., ed. edition.
- Chomsky, N. (1988). *Language and problems of knowledge: the Managua lectures*. Number 16 in Current studies in linguistics series. MIT Press, Cambridge, Mass.
- Clark, H. H. (2016). Depicting as a method of communication. *Psychological Review*, 123(3):324–347.
- Ebert, C. and Ebert, C. (2014). Gestures, Demonstratives, and the Attributive/Referential Distinction. Handout of a talk given at Semantics and Philosophy in Europe (SPE 7).
- Fox, D. (2007). Free Choice and the Theory of Scalar Implicatures. In Sauerland, U. and Stateva, P., editors, *Presupposition and Implicature in Compositional Semantics*, Palgrave

- Studies in Pragmatics, Language and Cognition, pages 71–120. Palgrave Macmillan UK, London.
- Fox, D. (2013). Presupposition projection from quantificational sentences: trivalence, local accommodation, and presupposition strengthening. *MIT web domain*.
- Fox, D. and Katzir, R. (2011). On the characterization of alternatives. *Natural Language Semantics*, 19(1):87–107.
- Gajewski, J. (2005). *Neg-raising: Polarity and Presupposition*. MIT dissertation. PhD thesis, MIT.
- George, B. R. (2008). A New Predictive Theory of Presupposition Projection. *Semantics and Linguistic Theory*, 18(0):358–375.
- Geurts, B. (1998). The Mechanisms of Denial. *Language*, 74(2):274–307.
- Geurts, B. (1999). *Presuppositions and pronouns*. Number v. 3 in Current research in the semantics/pragmatics interface. Elsevier, Amsterdam ; New York, 1st ed edition.
- Geurts, B. (2010). *Quantity Implicatures*. Cambridge University Press, Cambridge.
- Geurts, B. and van Tiel, B. (2016). When “All the Five Circles” are Four: New Exercises in Domain Restriction. *Topoi*, 35(1):109–122.
- Goldin-Meadow, S. and Brentari, D. (2017). Gesture, sign, and language: The coming of age of sign language and gesture studies. *Behavioral and Brain Sciences*, 40:e46.
- Grice, H. (1975). Logic and conversation. In *Syntax and Semantics 3: Speech Acts*, pages 41 – 58. Peter cole and jerry morgan edition.
- Heim, I. (1988). *The semantics of definite and indefinite noun phrases*. Outstanding dissertations in linguistics. Garland Pub, New York.
- Heim, I. (1990). Presupposition Projection. In *Presupposition, Lexical Meaning and Discourse Processes: Workshop Reader*, page 33, University of Nijmegen. R. van der Sandt.
- Heim, I. (1992). Presupposition Projection and the Semantics of Attitude Verbs. *Journal of Semantics*, 9(3):183–221.
- Heim, I. (2002). On the Projection Problem for Presuppositions. In *Formal Semantics*, pages 249–260. John Wiley & Sons, Ltd.
- Heim, I. and Kratzer, A. (1998). *Semantics in generative grammar*. Number 13 in Blackwell textbooks in linguistics. Blackwell, Malden, MA.
- Horn, L. (1972). *On the semantic properties of the logical operators in English*. Ph.D. thesis. PhD thesis, University of California at Los Angeles.
- Kadmon, N. (2001). *Formal pragmatics: semantics, pragmatics, presupposition, and focus*. Blackwell, Malden, Mass., 1. publ edition.
- Katzir, R. (2007). Structurally-defined alternatives. *Linguistics and Philosophy*, 30(6):669–690.
- Križ, M. (2015). *Aspects of homogeneity in the semantics of natural language*. PhD Thesis, University of Vienna.
- Križ, M. (2016). Homogeneity, Non-Maximality, and *all*. *Journal of Semantics*, 33(3):493–539.
- Križ, M. (2019). Homogeneity effects in natural language semantics. *Language and Linguistics Compass*, 13(11).
- Križ, M. and Spector, B. (2020). Interpreting plural predication: homogeneity and non-maximality. *Linguistics and Philosophy*.

- Levinson, S. C. (2000). *Presumptive meanings: the theory of generalized conversational implicature*. Language, speech, and communication. MIT Press, Cambridge, Mass.
- Löbner, S. (2000). Polarity in natural language: predication, quantification and negation in particular and characterizing sentences. *Linguistics and Philosophy*, 23(3):213–308.
- Mandelkern, M. (2016). A note on the architecture of presupposition. *Semantics and Pragmatics*, 9(0):13–1–24.
- Mayr, C. and Sauerland, U. (2016). Accommodation and the strongest meaning hypothesis. *Preproceedings of the Amsterdam Colloquium*.
- McNeill, D. (2005). *Gesture and thought*. University of Chicago Press, Chicago.
- Pfau, R. and Steinbach, M. (2006). Pluralization in sign and in speech: A cross-modal typological study. *Linguistic Typology*, 10(2).
- Potts, C. (2004). *The Logic of Conventional Implicatures*. Oxford University Press.
- Potts, C. (2015). Presupposition and Implicature. In *The Handbook of Contemporary Semantic Theory*, pages 168–202. John Wiley & Sons, Ltd.
- Sauerland, U. (2004). Scalar Implicatures in Complex Sentences. *Linguistics and Philosophy*, 27(3):367–391.
- Sauerland, U. (2012). The Computation of Scalar Implicatures: Pragmatic, Lexical or Grammatical?: Computation of Scalar Implicatures. *Language and Linguistics Compass*, 6(1):36–49.
- Schlenker, P. (2008). Presupposition Projection: the New Debate. *Semantics and Linguistic Theory*, 18(0):655–693.
- Schlenker, P. (2010). Presuppositions and Local Contexts. *Mind*, 119(474):377–391.
- Schlenker, P. (2016). The semantics–pragmatics interface. In Aloni, M. and Dekker, P., editors, *The Cambridge Handbook of Formal Semantics*, pages 664–727. Cambridge University Press, Cambridge.
- Schlenker, P. (2017). Outline of Music Semantics. *Music Perception: An Interdisciplinary Journal*, 35(1):3–37.
- Schlenker, P. (2018a). Gesture projection and cosuppositions. *Linguistics and Philosophy*, 41(3):295–365.
- Schlenker, P. (2018b). Iconic pragmatics. *Natural Language & Linguistic Theory*, 36(3):877–936.
- Schlenker, P. (2019a). Gestural semantics: Replicating the typology of linguistic inferences with pro- and post-speech gestures. *Natural Language & Linguistic Theory*, 37(2):735–784.
- Schlenker, P. (2019b). Prolegomena to Music Semantics. *Review of Philosophy and Psychology*, 10(1):35–111.
- Schlenker, P. (2020). Gestural grammar. *Natural Language & Linguistic Theory*, 38(3):887–936.
- Schlenker, P. (2021). Triggering Presuppositions. *Glossa: a journal of general linguistics*, 6(1).
- Simons, M., Tonhauser, J., Beaver, D., and Roberts, C. (2010). What projects and why. *Semantics and Linguistic Theory*, 20:309.
- Spector, B. (2007a). Aspects of the Pragmatics of Plural Morphology: On Higher-Order Implicatures. In Sauerland, U. and Stateva, P., editors, *Presupposition and Implicature in Compositional Semantics*, pages 243–281. Palgrave Macmillan UK, London.

- Spector, B. (2007b). Scalar Implicatures: Exhaustivity and Gricean Reasoning. In Aloni, M., Butler, A., and Dekker, P., editors, *Questions in Dynamic Semantics*, pages 225–249. BRILL.
- Spector, B. (2013). Homogeneity and plurals: From the strongest meaning hypothesis to supervaluations.
- Stalnaker, R. (1974). Pragmatic Presuppositions. In Stalnaker, R., editor, *Context and Content*, pages 47–62. Oxford University Press.
- Stalnaker, R. (1975). Presuppositions. In Hockney, D., Harper, W., and Freed, B., editors, *Contemporary Research in Philosophical Logic and Linguistic Semantics*, pages 31–41. Springer Netherlands, Dordrecht.
- Sudo, Y., Romoli, J., Hackl, M., and Fox, D. (2012). Presupposition Projection Out of Quantified Sentences: Strengthening, Local Accommodation and Inter-speaker Variation. In Aloni, M., Kimmelman, V., Roelofsen, F., Sassoon, G. W., Schulz, K., and Westera, M., editors, *Logic, Language and Meaning*, Lecture Notes in Computer Science, pages 210–219, Berlin, Heidelberg. Springer.
- Team, R. C. (2016). R: A Language and Environment for Statistical Computing (R Foundation for Statistical Computing, Vienna), Version 3.4.3.
- Tieu, L., Pasternak, R., Schlenker, P., and Chemla, E. (2018). Co-speech gesture projection: Evidence from inferential judgments. *Glossa: a journal of general linguistics*, 3(1):109.
- Tieu, L., Schlenker, P., and Chemla, E. (2019). Linguistic inferences without words. *Proceedings of the National Academy of Sciences*, 116(20):9796–9801.
- Tonhauser, J., Beaver, D., Roberts, C., and Simons, M. (2013). Toward a Taxonomy of Projective Content. *Language*, 89(1):66-109.
- Van Der Sandt, R. A. (1992). Presupposition Projection as Anaphora Resolution. *Journal of Semantics*, 9(4):333–377.
- van Rooij, R. and Schulz, K. (2004). Exhaustive Interpretation of Complex Sentences. *Journal of Logic, Language and Information*, 13(4):491–519.
- Zehr, J., Bill, C., Lyn, T., Jacopo, R., and Florian, S. (2016). Presupposition projection from the scope of None: Universal, existential, or both? *Semantics and Linguistic Theory*, 26(0):754–774.

Appendix I: Musical gestures and vocal gestures

As mentioned in section 3.3, we ran a parallel experiment on a different pool of participants using pro-speech onomatopoeias, i.e. iconic vocal sounds replacing one or several words in sentences, that we call ‘vocal gestures’, instead of musical gestures. The same four inference types were tested: scalar implicatures, presuppositions, supplements and homogeneity inferences, using paradigms and stimuli that were either perfectly analogous to the ones described throughout this paper, or very slightly different when vocal counterparts to musical stimuli could not be found. For instance, wherever an upward scale was used in our stimuli to evoke a rise in space to test for presuppositions, we used a whistle rising in frequency, and wherever a choir singing the French national anthem was used to evoke the action of singing to test for homogeneity inferences, we used a similar stimulus where the very same song was whistled instead of sung. Wherever a drum was used to evoke someone boxing to test for scalar implicatures, we used the onomatopoeia BOOM vocally pronounced.

We found a difference in the results collected from the experiment on musical gestures and the one on vocal gestures, which is surprising as most paradigms were perfectly symmetric. However, this experiment was ran on a smaller pool than the musical gestures experiment, and the lack of systematically significant effects of the inference type might thus be explained by a lack of power. We found indeed that the differences in the endorsement rates across both experiments were mainly not or marginally significant themselves, i.e. the responses given for the experiment on vocal gestures were not significantly different from the ones given for the experiment on musical gestures. The table below summarizes the comparisons of the distribution of the data between both experiments. For each inference type, we computed the interaction between inference type and the *Experiment* factor, whose two levels corresponded to the two experiments.¹⁸

Inference type	Environment	<i>p</i> -value for the [Inference type x Experiment] interaction	Significance interpretation
Scalar implicatures	Positive	0.74	No interaction
	Negative	0.08	Marginally significant interaction
Presuppositions	Question	0.61	No interaction
	None	0.43	No interaction
Supplements		< 0.001	Significant interaction (as expected)
Homogeneity	NP	0.05	Low interaction
	Predicate	0.29	No interaction

Figure 7: Comparison between musical gestures and vocal gestures experiments. The table displays the figures assessing the significance of the difference in distribution of the data collected across experiments. None of the data subsets were significantly different across both experiments, except for supplements, for which we found no contrast in endorsement due to an unexpected interpretation of the vocal stimulus.

¹⁸The detailed analyses and statistical scripts to compute interactions are available in the RESULTS folder at https://osf.io/hw45u/?view_only=89f983db777f49e9a6f5b41b3dea60d6

Appendix II: Statistical models

In this appendix, we provide the generalized linear mixed-effects models we used to analyze the data. They were all inspired by that of (Tieu et al., 2019), whose analysis script is open source. For each model, we justify the structure of the model and in particular its random effects structure by the design (as recommended in Barr et al. (2013)). We also provide the coefficients found for each model. For the sake of clarity, we report the statistical models in a raw fashion (i.e. the R code lines). Useful information about the syntax of R, which can be helpful to read these formulas, can be found online and are accessible at this link: https://github.com/clayford/LMEMInR/blob/master/lme4_cheat_sheet.Rmd.

Scalar implicatures

In the scalar implicatures paradigms, two main factors could have had an effect on the endorsement of the target inferences: the gesture factor (`GestureC`), which two levels correspond to the target and control premises contrasting two alternative realizations of the drum sound mimicking someone boxing, `DRUM×1` and `DRUM×2`; and the inference factor (`InferenceC`), which two levels correspond to the target and baseline inferences contrasting two possible interpretations of the the drum sounds (‘some’ vs ‘a lot’ in the positive environment; ‘some’ vs ‘none’ in the negative environment).

Here, we were interested in the `GestureC * InferenceC` interaction to ensure that the contrast found for the target premise was not due to a default bias in endorsement for one inference over the other, by comparing the difference between the endorsement of the target and that of its negation for the target premise and the same difference for the control premise. We did *not* include the environment factor in the model (positive *vs* negative) because there was no theoretical reason to expect a difference in these interactions between the positive and the negative environment. `GestureC`, `InferenceC`, and their interaction were thus used as fixed effects in our model, while this interaction *by subject* was used as a random effect, accounting for the variability across participants in (i) the interpretation of the premise, (ii) the interpretation of the inference, and (iii) the interaction between (i) and (ii).

Model

```
value ~ GestureC * InferenceC + (1 + GestureC * InferenceC|SubjID)
```

This first model failed to converge. We followed the same procedure as (Tieu et al., 2019) in simplifying the random effects structure, and removed the interaction between the two factors `GestureC` and `InferenceC` from the random effects as follows:

Simplification of random effects structure (as in (Tieu et al., 2019))

```
value ~ GestureC * InferenceC + (1 + GestureC + InferenceC|SubjID)
```

Environment	(Intercept)	GestureC	InferenceC	GestureC:InferenceC
Positive	52.60377	-0.79245	-0.07547	116.60377
Negative	57.005	1.557	7.575	44.698

Presuppositions

There was no theoretical reason for predicting an interaction in presuppositional behavior between question formation and projection under ‘none’, as the two projection tests are independent. Inferences from questions and from ‘none’ were thus analyzed separately. We are rather interested in the effect of the inference type (presupposition vs no presupposition) on the responses.

To ensure that the inference containing presupposition was not due to a by-default preference for this kind of inference, the endorsement of the presupposition was contrasted with that of its negation (baseline inference), as a control. The model thus used the inference factor `InferenceC`, which levels correspond to the presupposition/no presupposition, as a fixed effect, and this same factor *by subject* was used as a random effect, to capture the possibility that each participants may simply tend to endorse presuppositions differently.

Model

```
value ~ InferenceC + (1 + InferenceC | SubjID)
```

This first model failed to converge, leading us to simplify the random effect structure by only keeping the `SubjID` factor as a random effect, to capture the variability in intercept across participants:

Simplification of random effects structure

```
value ~ InferenceC + (1 | SubjID )
```

Environment	(Intercept)	InferenceC
Question	49.48	-15.60
‘none’	51.65	2.74

Supplements

The case for supplements is symmetric to the paradigm for scalar implicatures in 6, where both the gesture factor `GestureC`, describing the two types of premise (the target premise where the musical gesture is expected to be interpreted as a non-restrictive relative clause, and the control premise where the musical gesture is made at-issue by using ‘like this’) and the inference factor `InferenceC`, describing the two types of inference (supplemental or not) were used as fixed effects in the model, while the interaction *by participant* was used as a random effect in the maximal model below:

Here, we are interested in the interaction between the `Gesture` factor, which two levels represent the two forms of sentences (with a pro-speech gesture standing for a non-restrictive relative clause, and with the deictic ‘like this’) and the `Inference` factor, which two levels correspond to the supplemental inference (if X, then X would have happened in a certain way) and the inference without the supplemental information (if X, then X would not have necessarily happened in this same way). As the control inference using ‘not necessarily’ was not the exact negation of the target inference for reasons of simplicity, we had no strong prediction as to how differently the target and control inference would be endorsed for the control premise using ‘like this’; but we expected this difference to be important for the target inference using a post-speech musical gesture (without ‘like

this’) which was expected to trigger a supplemental inference just as post-speech gestures do (Tieu et al., 2019; Schlenker, 2018a). We thus expected an interaction between the two factors, with a higher difference in endorsement between the target inference and the baseline inference in response to the target premise than in response to the control premise.

Model

value ~ GestureC * InferenceC + (1 + GestureC * InferenceC | SubjID)

This first model failed to converge, leading us to simplify as before the random effect structure by removing the interaction:

Simplification of random effects structure 1

value ~ GestureC * InferenceC + (1 + GestureC + InferenceC | SubjID)

This second model converged but did not allow for interaction testing (removing the interaction from the model prevented the model from converging), leading us to simplify the random effect structure even more:

Simplification of random effects structure 2

value ~ GestureC * InferenceC + (1 + InferenceC | SubjID)

(Intercept)	GestureC	InferenceC	GestureC:InferenceC
57.175	-7.085	-26.858	8.774

Homogeneity inferences - NP and Predicate

The case for homogeneity is analogous to that of presuppositions in 6, where only inference type was used as a fixed effect in the model, while the effect of inference type **by participant** (testing whether each participant had a personal tendency to endorse the inferences) was used as a random effect. As we were interested in whether the homogeneous (‘all’ or ‘none’) inference would be preferred to the non-homogeneous one in each environment (positive and negative), we are here only interested in the effect of inference type (homogeneous *vs* non-homogeneous).

Model

value ~ InferenceTypeC + (1 + InferenceTypeC | SubjID)

This model did not converge, so we decided, as with presuppositions, to go with the most minimal random effect structure only accounting for the difference in intercepts between participants:

Simplification of random effects structure

value ~ InferenceTypeC + (1 | SubjID)

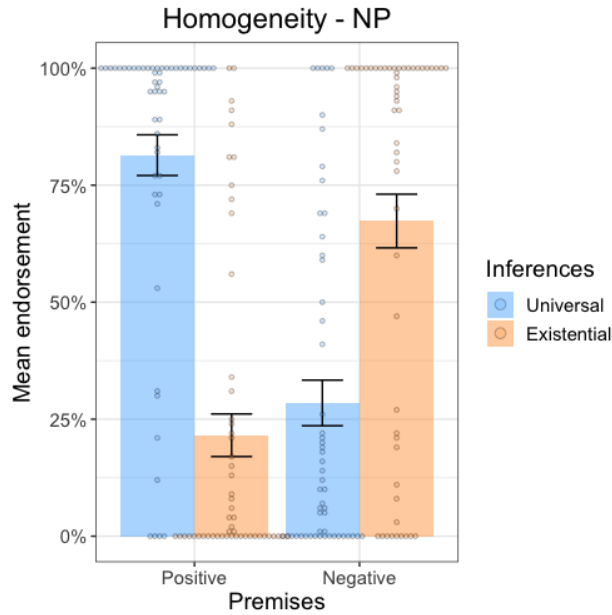


Figure 8: Reading flipping under negation

Hypothesis tested	Environment	(Intercept)	InferenceC
NP	Positive	51.49	-59.85
NP	Negative	47.91	-38.87
Predicate	Positive	54.71	-61.34
Predicate	Negative	40.31	-37.34

Note that in our analysis, we were interested in inference type, i.e. whether the homogeneous inference was significantly more endorsed than its non-homogeneous counterpart. However, as the editor pointed out, it would also have made sense to analyze the inference type in terms of *universal* or *existential*. As described in the paradigm in 3.4, we expected musical gestures (that were punctuated repetitions) to have a universal reading in the positive environment (‘all harps’) and to get an existential reading in the negative environment (‘at least one harp’, i.e. ‘some harps’), just as the reading of gestures and visual animations were shown to flip from universal to existential depending on the context in (Tieu et al., 2019). If we perform such an analysis, we are interested in the interaction between environment (positive *vs* negative) and inference type (understood as universal vs existential, and not as homogeneous vs non-homogeneous anymore). We plot the results with this new analysis and report the significance and coefficients of what the model would have been in this situation below:

Model

```
value ~ InferenceTypeC * Environment + (1 + InferenceTypeC * Environment | SubjID)
```

This model failed to converge, leading to a simplification of the random effects structure as follows:

Simplified model

```
value ~ InferenceTypeC * Environment + (1 + InferenceTypeC + Environment | SubjID)
```

The comparison of this model to the same model without the interaction between inference type and environment showed a significant difference between the two models ($\chi^2 = 84, p < 0.001$). This supports the idea that the reading of a same musical gesture flips under negation, which is consistent with our first analysis.

Iconicity controls

For each iconicity control, we were interested in the difference in endorsement between the inference containing the matching interpretation of the iconic modulation and the inference that contained the opposite non-matching interpretation. For instance, we wanted to know whether the inference containing the interpretation of a long upward scale as a high mountain was more endorsed than the inference containing the interpretation of a long upward scale as a low mountain. We thus counted inference type as a fixed effect. To model the possible differences in the tendencies to endorse both types of inferences in a certain way for each participant, we included inference type by subject as a random effect in the model, as shown:

Model

$\text{value} \sim \text{InferenceTypeC} + (1 + \text{InferenceTypeC} | \text{SubjID})$

Control	(Intercept)	InferenceC
1	55.09	52.11
2	52.75	-28.37

Appendix III: Inferential mechanisms

Here, we provide a loose description of the underlying mechanisms responsible for the four types of inferences tested. Although an important part of its content is itself subject to debate, the aim of this table is merely to provide some very basic analysis of each inference to facilitate the reading of each section; it does *not* aim at exhaustively accounting for all possible theoretical models.

Inference	Input	Mechanism	Output
PRESUPPOSITION	Bivalent ²² truth-conditions (from either linguistic or non-linguistic items)	Takes the input q and produces a pair $\langle p, p' \rangle$ with q equivalent to $p \ \& \ p'$ where p is the presupposition.	Presupposition + Assertion ²³
SCALAR IMPLICATURE	Alternatives (non-necessary linguistic) with their bivalent content.	Competition between alternatives (Gricean or grammatical mechanism)	Ordinary meaning + Strengthened meaning
SUPPLEMENT	Bivalent content + syntactic form	Whatever allows a non-restrictive relative clause to get the meaning of a supplement [e.g. comma intonation (Potts, 2015)]	Supplemental inference
HOMOGENEITY INFERENCE	<p><i>Previous theories:</i> Definite plural NP [or non-linguistic item referring to definite pluralities such as punctuated gestures or sounds]</p> <p>----- <i>More recently</i> (Križ, 2020): Trivalent predicate [or non-linguistic predicate like gestural or musical]</p>	<p>Assigns the input a universal reading ('all') in positive environments and an existential reading in negative environments</p> <p>----- If a group is in the extension of the predicate, then all members of the group are in the extension of the predicate as well (positive environment) or none of them (negative environments)</p>	Homogeneous inference

Figure 9: Description of each inferential type

²² Bivalence refers to the existence of pre-conditions to the meaning.

²³ The output is actually closer to \langle presupposition, presupposition+assertion \rangle , because the assertion in \langle presupposition, assertion \rangle is underdetermined.

Chapter 3

Meaning from music in audiovisual scenes

Purpose

The previous chapter explored the interaction between music and language, and revealed that musical meaning can be integrated with linguistic meaning when musical stimuli are embedded in language. In this chapter, we explore musical meaning through its interaction with visual animations. By creating audiovisual stimuli in which musical meaning combines with visual information, we show that musical properties can be mapped onto properties of visual objects through systematic rules.

HOW MUSIC CAN REFER TO OBJECTS

COREFERENCE FROM DYNAMICS, PITCH AND TIMBRE IN AUDIOVISUAL STIMULI *

Léo Migotti

Institut Jean Nicod (ENS-EHESS-CNRS)

Emmanuel Chemla

Département d'Etudes Cognitives (ENS-EHESS-CNRS)
Laboratoire de Sciences Cognitives et Psycholinguistique (CNRS)

Philippe Schlenker

Institut Jean Nicod (ENS-EHESS-CNRS)
Department of Linguistics, New York University

ABSTRACT

Building upon recent theories of musical meaning that rely on the assumption that music conveys information about an extra-musical world, this paper presents an experimental paradigm testing semantic associations between musical properties and visual objects. It first brings evidence that when an object is present in a visual scene paired with some music, pitch and dynamics can be interpreted as the level of energy of this object, or as its distance from a viewpoint. Second, it shows that pitch and dynamics can be associated with specific objects in scenes that contain multiple objects. Third, it shows that timbre too encodes reference to objects. Together, these results demonstrate that musical properties establish reference to specific objects, and that those reference relations are established between music and objects themselves, rather than between music and the scenes containing these objects.

Keywords Music semantics | Coreference | Pitch | Dynamics | Timbre | Audiovisual integration

* *Contributions:* LM, EC and PS designed the experiment based on previous work by PS. LM and EC worked on the details of the experimental paradigm, and LM and PS worked on clarifying the underlying theories. LM created the audiovisual stimuli, programmed the experiment, managed data collection, analyzed the data and wrote the paper, with contributions and proofreading by PS and EC.

Contents

1	Introduction	86
2	Background	87
2.1	Variables across media	88
2.2	Coreference in music	89
3	Testing the semantics of pitch and dynamics	91
3.1	Theoretical framework	91
3.2	Experimental design	92
3.2.1	General experimental procedure	92
3.2.2	Design and stimuli of block 1	93
3.3	Results	95
4	Testing coreference with pitch and dynamics	97
4.1	Experimental paradigm	97
4.1.1	Target stimuli	97
4.1.2	Control stimuli	98
4.2	Results	99
5	Testing coreference with timbre	102
5.1	Motivation	102
5.2	Experimental design	104
5.3	Results	105
6	Conclusion	106
	Bibliography	111

1 Introduction

Although many accounts of musical meaning have highlighted the ability of music to represent things in the non-musical world,² only recent work has established formal theories explaining how such mechanisms are possible (Schlenker, 2017). This research agenda relies on the idea that music can be *true* of certain situations: there are certain conditions under which a given musical excerpt can represent (or *denote*) a given situation, and systematic rules that map a certain musical form to all things it can be true of.

Among all candidate musical properties involved in these mechanisms, some hypotheses have been given in the literature about the interpretation of pitch and dynamics (Schlenker, 2017; Eitan and Granot, 2006). In particular, pitch is claimed to be interpreted as the level of energy of those objects, while dynamics are claimed to be either interpreted as energy or distance from a viewpoint. For a situation to be true of a rise in pitch for instance, it requires that this situation involve a gain in energy.

But music can do more than just represent some situations better than others. It has been highlighted that it can also refer to objects in these situations (Schlenker, 2022): not only are musical properties interpreted when objects are given, they might also help a listener understand that a certain musical passage is *about* (or refers to) a certain object, and that another passage refers to another object. For instance, a recurring melodic pattern might systematically refer to a same character in opera and hence generate **coreference**. By contrast, certain changes in dynamics (i.e. playing more or less loudly) might instead indicate that two objects are being referred to instead of one, hence generating **disjoint reference**.

The same body of research has finally proposed that timbre is of a particular nature with regards to its semantic interpretation. Because timbre seems to have the inherent ability to

²Among many examples from the literature, we can cite (i) several accounts on how music can evoke movement in Godøy and Leman (2010), in particular thanks to pitch height and loudness Eitan (2013); Eitan and Granot (2004) how some musical properties can be interpreted as analogs to physical forces in Larson (2012), and more theoretical accounts on the music/movement relationship in Clarke (2001). (ii) A rich literature on how music can convey emotional information includes Blumstein et al. (2012), Juslin and Laukka (2003) on similarities between musical and vocal performance in their ability to trigger emotional effects, and Koelsch (2012) for a review in brain-related processes responsible for these effects. A substantial body of research has also investigated how music and language might share some properties explaining their ability to convey meaning in Cross and Woodruff (2009), and how different parameters such as pitch height or intensity can related to different emotional interpretations in music and speech. (iv) Sievers et al. (2013, 2019) offer an account on how musical and visual properties can encode similar emotional effects as well. See also (v) Gabrielsson and Lindström (2010) for a discussion of the contribution of uninterpreted musical structure. Schlenker (2017) provides further references in particular to review works on related topics.

be indicative of the very nature of an object, it often binds musical motives to objects in a direct fashion. Schlenker (2022) argues that orchestration makes great use of the semantics of timbre to indicate coreference or disjoint reference in orchestra pieces. To give listeners the impression that two repetitions of a same musical motive correspond to two different objects, composers might typically vary timbre.

This paper has three main goals. First, it tests the hypotheses about the semantic interpretation of pitch and dynamics through an audiovisual experimental paradigm involving musical sounds and objects in visual animations. Second, it tests whether pitch and dynamics can make reference to specific objects, i.e. whether it is possible to manipulate pitch and dynamics so that one knows from the music which object is being referred to. Third, it tests the semantics of timbre and shows that timbre can establish reference to objects, and that these reference relations are established to objects directly, and not merely to the general scene(s) involving these objects. After presenting some background about coreferential effects across media and in music more specifically in section 2, we present the experimental paradigm in three steps. Section 3 presents how the predictions on the semantics of pitch and dynamics were tested. Section 4 explains how those results were reused to establish the ability of pitch and dynamics to make reference to objects. Section 5 finally presents how timbre too can make reference to objects and provides an argument in favor of a direct association between musical property and object: listeners bind music to objects and not merely to visual scenes.

2 Background

While recent work has established that referring to external objects is not unique to language, most accounts of reference relations in the literature address the case of natural languages. The idea that it is possible for expressions in natural languages to refer to a same entity (establishing coreference to this entity) is an old one. Consider the following sentence in (1).

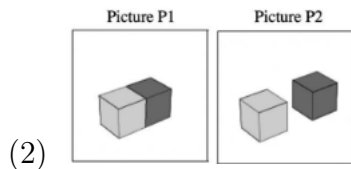
- (1) Mary said that she would come.
 - 1a. Mary_i said that she_i would come.
 - 1b. Mary_i said that she_j would come.

The sentence in (1) can be interpreted as either 1a, namely that Mary said that Mary herself would come, or else as 1b, namely that Mary said that another person referred to by the third-person pronoun ‘she’ would come. In 1a., ‘Mary’ and ‘she’ are coreferential: they refer to the same person Mary, while in 1b., the relation between ‘Mary’ and ‘she’ is that of a disjoint reference: ‘Mary’ and ‘she’ refer to two different persons. Note that one way of explaining how the ambiguity is lifted (which can be traced back to Montague (1973) and Karttunen (1969)) is to call for a notion of variable (aka discourse referent), i.e. abstract objects assigned to different items in the sentence, as shown in (1) through indexes i and j .

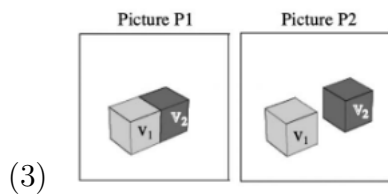
Recent work has evidenced the existence of coreference relations beyond natural language. The next section highlights two particular instances in pictorial narratives and dance.

2.1 Variables across media

In Abusch (2013, 2017, 2020), coreference is found in comics where the depiction of a character shown on two different panels is understood as depicting the same character. The same idea can be represented in (2) taken from Abusch (2017) and represented in Schlenker et al. (2023) as follows:



The most salient interpretation of (2) is that the light cube in Picture 2 refers to the same cube as the light cube in Picture 1, and that the dark cube in Picture 2 refers to the same cube as the dark cube in Picture 1. Another (less natural) interpretation of (2) is that instead, the dark cube in Picture 1 was replaced by another dark identical dark cube in Picture 2. Although this sequence of pictures is thus initially ambiguous, the former coreferential interpretation seems preferred. And here again, one possible way to account for these coreference relations is to posit the existence of two discourse referents that index each cube across the two panels, as shown in (3).



Because these analyses of coreference relations arise both in concrete visual representations such as comics but also in abstract pictorial narratives such as that in (2), recent work has investigated whether coreference can be found in other visual representations beyond pictures. In particular, Patel-Grosz et al. (2018) has argued that there is coreference and disjoint reference in Bharatanatyam, a narrative Indian dance form. The authors found that reference to a same character involved in two different activities was marked by a fluid transition between the two positions referring to those activities. By contrast, disjoint reference was marked by a more structured sequence of intermediate dance positions marking the difference from the referent of the first dance position to the referent of the second dance position. Coreference seems, therefore, also visually encoded in abstract forms such as dance.

2.2 Coreference in music

If coreference relations are not exclusive to natural languages but also exist in sequences of pictures and sequences of dance positions, it becomes plausible that such relations can also be expressed beyond the visual modality, and in particular in audition.

Several examples from Schlenker (2022) suggest that genuine referential ambiguities can be found in music, and that conveying coreference can be achieved through different means, notably through orchestrations of pieces originally written for a single instrument, and through choreographies. For instance, in Chopin's Mazurka op. 33 n°2,³ a same melodic pattern of the form AB is repeated in A'B' with $A = A'$ and $B = B'$.

Based solely on this structure, and under the hypothesis that two objects are here represented,⁴ at least two interpretations are possible: one object corresponds to AB and another one to A'B'; or one object corresponds to A and A', and another one to B and B'. But once one notices Chopin's dynamics indication on the score (circled in red) that A and B

³See for instance [this interpretation](#) by Arthur Rubinstein

⁴As noted by Schlenker, this need not be the case.

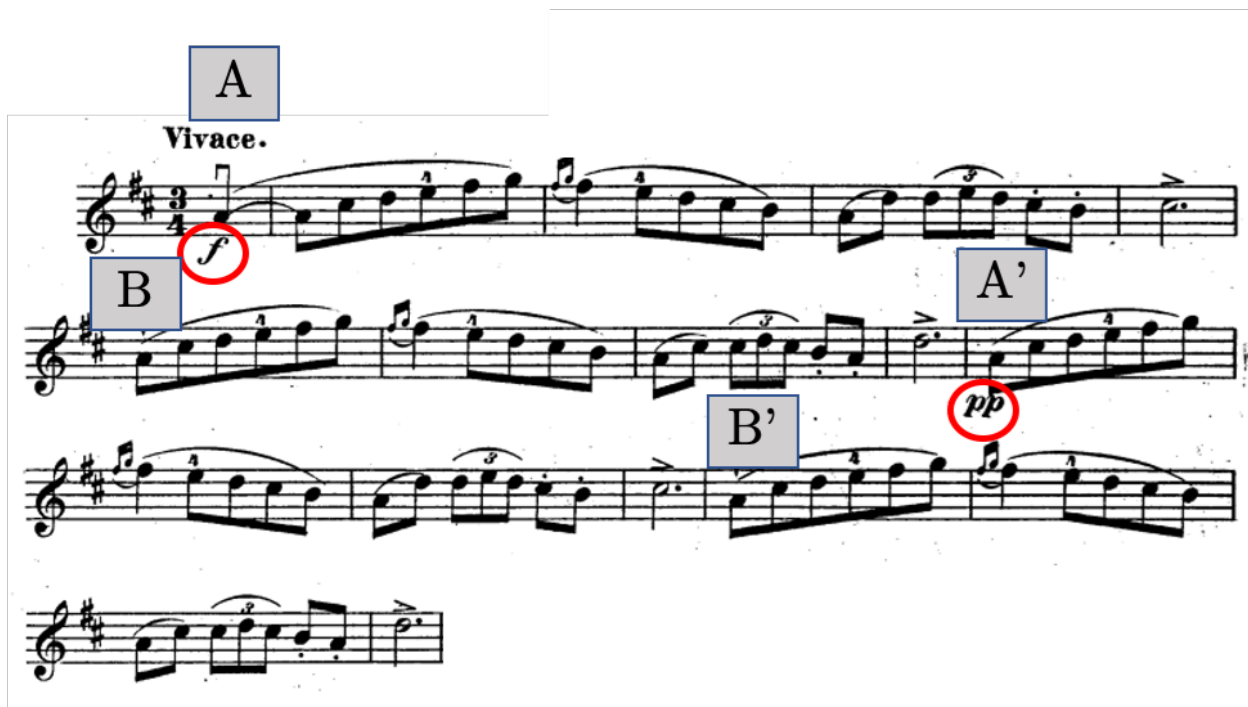


Figure 1: First bars of Chopin's mazurka Op. 33 n°2
Simplified version taken from the violin part of the arrangement for piano and violin by G. Adolfson (1870)
Retrieved from imslp.org

have to be played (*forte*) (i.e. quite loud) and that A' and B' have to be played (*pianissimo*) (i.e. very soft), the second interpretation becomes unlikely.

In many orchestrations of this piece,⁵ AB and A'B' are assigned different groups of musical instruments; hence each passage is heard with a different timbre, reinforcing the disjoint reference between AB and A'B'. As discussed in Schlenker (2022), this disjoint reference was also made salient by visual means in Russian choreographer Michel Fokine's choreography on an orchestration of Chopin's Mazurka named *Les Sylphides* in which parts AB and A'B' involve different groups of dancers.

Together, these cases suggest that reference to objects can be found in music. After establishing how pitch and dynamics are semantically interpreted when paired to objects in section 3, we will then explore whether they can trigger referential effects in section 4 before introducing the special case of timbre in section 5.

⁵See Schlenker (2022) for a list of works involving similar orchestration choices, among which Benjamin Britten's, Roy Douglas's, Maurice Keller's, and Gordon Jacob's.

3 Testing the semantics of pitch and dynamics

This section explores whether hypotheses about the semantic interpretation of pitch and dynamics as energy levels and distance of objects as posited in Schlenker (2017) can be experimentally confirmed. We present these hypotheses in the next subsections.

3.1 Theoretical framework

Dynamics. In Schlenker’s model, dynamics (referred to as ‘loudness’ in the work cited⁶) are argued to be an ambiguous musical property. Dynamics can either be interpreted as distance to an object from a perspectival center (**the louder the music, the closer the object**), or about its level of energy (**the louder the music, the more energy the object has**). These two assumptions are argued to come from normal (i.e. non-musical) auditory cognition. In our environment, a sound produced by an object is naturally perceived as louder when the object gets closer; and sound-producing objects that have high energy levels tend to produce louder sounds. In the first block of our experiment, we tested whether dynamics could indeed be semantically interpreted as either energy or distance.

Pitch. In the same model, when the nature of an object denoted by music is given, pitch is interpreted as the level of energy of that object: **the higher the pitch, the higher the level of energy of the denoted object**. Such a semantics for pitch is argued to be lifted from normal auditory cognition as well: generally speaking, things tend to produce higher-pitched or higher-frequency sounds when excited (Schlenker, 2017). Similarly, animal vocalizations and human voice seem to rely on high pitch to convey high level of emotional arousal or energy (Bachorowski, J. A., 2008; Seyfarth and Cheney, 2003). Although many other hypotheses on the interpretation of pitch have been experimentally tested in the literature,⁷ this specific interpretation in terms of energy has not. Finally, besides this interpretation of pitch as energy, and in order to have a symmetric paradigm involving as many stimuli manipulating pitch as stimuli manipulating loudness, we ran an exploratory test of the hypothesis according to which pitch could also be interpreted as distance, though not backed up by previous

⁶We prefer the term ‘dynamics’ to that of ‘loudness’ to get closer to music terminology: although strictly speaking, the main property co-varying with dynamics in music is indeed loudness (or volume), dynamics might have extra-dimensions to them (Olsen, n.d.)

⁷For instance, Eitan (2013); Eitan and Granot (2004) on the mapping of pitch to a vertical position in space.

theoretical work. The rule tested here was that **the higher the pitch, the closer the object.**

Note that in both cases, and although these inferential mechanisms might have been lifted from normal auditory cognition, they seem to apply to silent objects just as much. For instance, we still expect a louder sound to represent a silent object closer to us compared to a softer one.

3.2 Experimental design

In this section, we first provide some experimental details that were common to all three blocks of the experiment, before digging into the design of the first one.

3.2.1 General experimental procedure

The elements mentioned in this subsection hold for all three blocks of the experiment, which was run on a pool of 60 participants recruited through the online platform Prolific[©]. All participants first went through a few slides presenting the experiment and testing the audio level to be fixed for the rest of the experiment. All of them gave online informed consent before starting. The experiment had three blocks, the first of which was always the block testing for the semantics of pitch and dynamics presented in this section. The order of the two subsequent blocks, testing for pitch and loudness, and for timbre respectively, was randomized.⁸ The experiment took between about 15 to 25 minutes to complete. Participants were paid through the platform at an hourly rate of \$9, based on an expected mean experimental time of 20 minutes.

An attention check was inserted after the experimental block 2 to make sure participants were still listening and focused. It consisted of a video picked at random among all stimuli with a voice instead of a musical stimulus indicating which response to give, and to type a given word in a text box appearing on the following screen. Among all 60 participants, 5 failed the attention check and were therefore removed from the data. This criterion, along with all hypotheses to be tested, the design of the experiment and the statistical analyses

⁸A post-hoc linear mixed-effect model revealed no significant effect of order on responses from block 2, be it for stimuli involving a change in dynamics ($\chi^2 = 0.4192, p = 0.5173$) or a change in pitch ($\chi^2 = 0.8188, p = 0.3655$)

scripts were preregistered prior to data collection. All files were archived as open data at [this link](#).⁹ The preregistration is available [here](#). In this paper, we present the three blocks separately for clarity.

All stimuli were created from a combination of visual and audio stimuli and available [here](#). Audio stimuli were generated with Garageband[©] (version 10.3.5). The corresponding Garageband folder containing information on loudness levels and changes, timbre, duration and tempo is accessible in the same folder. Visual stimuli were generated with Open source software Blender[©] (version 3.4.1 2022-12-20) using the scripting interface allowing for programming the visual animations and easily changing its main properties. Each script used to generate the visual animations and to incorporate the audio stimuli is available in the corresponding folders for each block of the experiment.¹⁰

The following subsection provides details on the design and stimuli from the first experimental block testing for the semantics of pitch and dynamics.

3.2.2 Design and stimuli of block 1

The purpose of the first block of the experiment was to test whether participants endorsed the expected semantic interpretations derived from hypotheses in section 3.1. To that end, we created audiovisual stimuli¹¹ in which six musical stimuli (a crescendo and a decrescendo - i.e. an increase and decrease in loudness, respectively -, a rise and a fall in pitch along a

⁹Any modification to the preregistered statistics, mainly pertaining to matters of (i) visualization, (ii) data presentation or (iii) the investigation of the statistical significance of some contrasts not predicted by the theory, which could not be anticipated from pilot experiments, are justified on the scripts.

¹⁰While all information is available in the aforementioned documents, we here provide justification for some of the choices made in the design of the stimuli.

(i) *On the audio stimuli*: For the two 1st blocks of the experiment, in which timbre had to be constant, we used the timbre ‘Soft Square Lead’ available in the Garageband instruments library, which sounds close to a pure sound although created through square waves instead of sine waves. Two main reasons motivated this choice. First, although initially, we were planning on using pure sounds, pilots on informants suggested that it seemed to alter the naturality of the task, as the pure sounds were distracting. Second, we still wanted to work with a rare synthesized timbre to avoid personal or cultural associations made from well-known instruments timbres, and to avoid trivial associations with the musical instruments themselves rather than with external objects. Because it is reasonable to posit that participants were either never or only very rarely exposed to that very specific timbre, the peripheral associations with it were arguably interacting with the task less.

(ii) *On the audiovisual combinations*: One will note that while our visual stimuli were continuous scenes, our audio stimuli were discrete: the changes in dynamics or in pitch occurred between discrete notes, except for the 3rd block of the experiment. The reason for this was that it is arguably a good approximation of how musical meaning works: music is indeed a sequence of discrete musical events, while the events in the world that it represents are continuous.

¹¹We were here inspired by several audiovisual tasks from previous research, which have been very fruitful to establish some relations between music and world properties in work such as Blumstein et al. (2012) where authors paired soundtracks with videos to test for the effect of musical non-linearities (such as abrupt frequency drops), or Sievers et al. (2019) who developed a computer program from which participants could use cursors to create a musical track from the video of a ball and *vice versa*.

	Dynamics		Pitch			
	Crescendo	Decrescendo	Ascending		Descending	
			Major	Minor	Major	Minor
Moving closer	[-]	[-]	[-]	[-]	[-]	[-]
Gaining energy	[-]	[-]	[-]	[-]	[-]	[-]
Moving away	[-]	[-]	[-]	[-]	[-]	[-]
Losing energy	[-]	[-]	[-]	[-]	[-]	[-]

 Congruent stimuli
 Incongruent stimuli

Table 1: Distribution of the stimuli across conditions in experimental block 1
(All stimuli can be accessed by clicking on the corresponding cell.)

scale in major mode and in minor mode) were paired with either a congruent or incongruent visual among four visual animations (an object moving away, moving closer, gaining energy or losing energy). For instance, under the assumption that dynamics are interpreted as distance, and given a crescendo, the congruent stimulus involves an object moving closer, and the incongruent version would involve an object moving away.

The congruent stimuli testing for loudness involved either (i) a crescendo paired with an object approaching or gaining energy, or (ii) a decrescendo paired with an object either moving away or losing energy. **The incongruent stimuli testing for loudness involved either** (i) a crescendo paired with an object moving away or losing energy, or (ii) a decrescendo paired with an object either moving closer or gaining energy. **The congruent stimuli testing for pitch** involved either (i) an ascending scale in major or minor mode paired with an object gaining energy, or (ii) a descending scale in major or minor mode paired with an object losing energy. **The incongruent stimuli testing for pitch** involved the same scales paired with a change in energy in the opposite direction. Energy gain/loss was visually indicated through a shaking movement of the target object with a linearly increasing/decreasing amplitude. Although the stimuli used for illustration here only involve cubes, each stimulus also had a cylinder version in order to make sure that the evidenced effects, if any, were not specific to a specific 3D shape and to introduce variability in the stimuli. A summary of the design with accessible stimuli is available in Table 1.

For each trial, participants were presented with a pair of videos containing the same musical sound. One video was congruent with the predicted interpretation, and one was

incongruent with that interpretation. Participants were then asked to click on a button under the one they preferred,¹² as shown in the screenshot of the experimental interface in Figure 2. The position of the congruent video on the screen was randomized. Trial order was fully randomized as well. The choice was binary so that participants had to choose one stimulus or the other.¹³ Our prediction was that the congruent video would be systematically preferred.

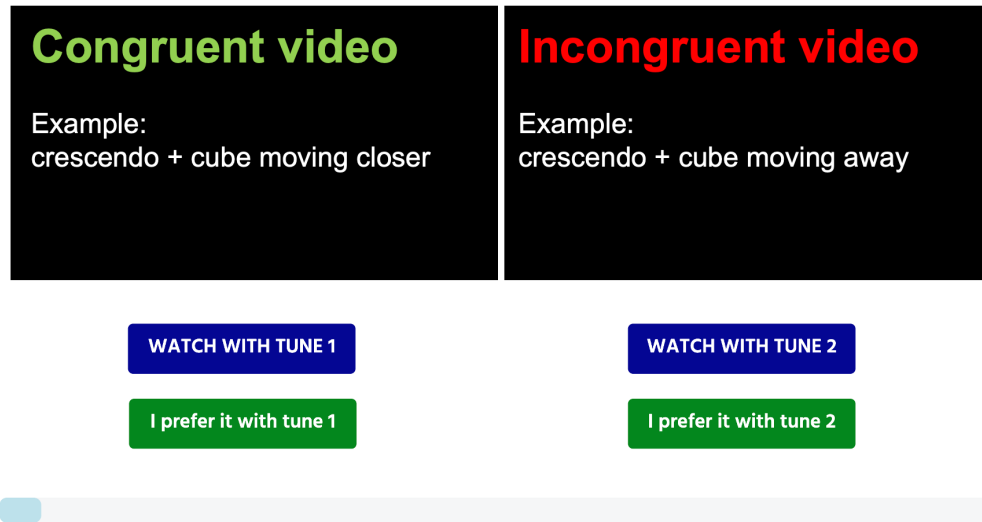


Figure 2: Experimental task - 1st experimental block

3.3 Results

The results in Figure 3 confirm that the predicted (congruent) interpretations of both pitch and dynamics were significantly preferred compared to incongruent interpretations.¹⁴ An increase (or decrease, respectively) in dynamics was systematically judged as being more

¹²For a review of methods and a justification of the use of preference tasks in linguistics, which are particularly relevant to highlight contrasts compared to interpretation judgment tasks, see Ionin and Zyzik (2014)

¹³While it could be argued that this could lead to possible biases, forcing a binary choice would in any case not show any preference for any of the two videos if there were none, as the choice would then be random.

¹⁴For each musical property (pitch and dynamics), and each interpretation (as energy level or distance), we wanted to check whether participants chose the congruent interpretation significantly more than the incongruent one, i.e. that for each participant, the proportion of congruent stimuli was significantly higher than 50%. We therefore ran 4 one-sample t-tests. (i) For the distance interpretation of loudness: $t = 4.3501, df = 104, p = 3.186e - 05$, (ii) for the energy interpretation of loudness, $t = 3.5158, df = 107, p - value = 0.0006442$, (iii) for the distance interpretation of pitch, $t = 2.1689, df = 213, p - value = 0.03119$ and (iv) for the energy interpretation of pitch, $t = 6.2841, df = 214, p - value = 1.82e - 09$. We also ran linear mixed-effects model on the data for both properties separately: (i) for loudness, neither object type ($\chi^2 = 0.368, p = 0.5441$) nor the visual interpretation ($\chi^2 = 0.368, p = 0.5441$) had a significant effect on the mean endorsement of congruent stimuli picked, and (ii) for pitch, neither object type ($\chi^2 = 0.8526, p = 0.3558$) nor mode (major or minor) ($\chi^2 = 2.2412, p = 0.3261$) had a significant effect on the responses. However, the visual interpretation had a significant effect ($\chi^2 = 6.6422, p = 0.009959$) on the responses, indicating that the interpretation of pitch as energy was more easily accessible to participants than that as distance.

congruent with an object moving closer (moving away, respectively), and of an increase (or decrease, respectively) in energy. This confirms that dynamics had indeed at least two available interpretations as both distance and energy level. Similarly, the results from pitch confirm our predictions: a rise in pitch (or fall, respectively) was systematically more associated with the object gaining energy (or losing energy, respectively). Neither the object type (cube/cylinder) nor the mode of the scale (major/minor) had any significant effect on the responses.

Interestingly, the exploratory test on the interpretation of pitch as distance, which was not backed up by theory, revealed a significant preference for a mapping from a rise in pitch to a decrease in object distance: the higher the pitch, the closer the associated object. Although this slight preference was found, the results finally show a significant interaction between the two available interpretations of pitch with a significantly higher endorsement rate for the interpretation as energy than for the interpretation as distance, suggesting that the interpretation of pitch as energy was more easily accessible.

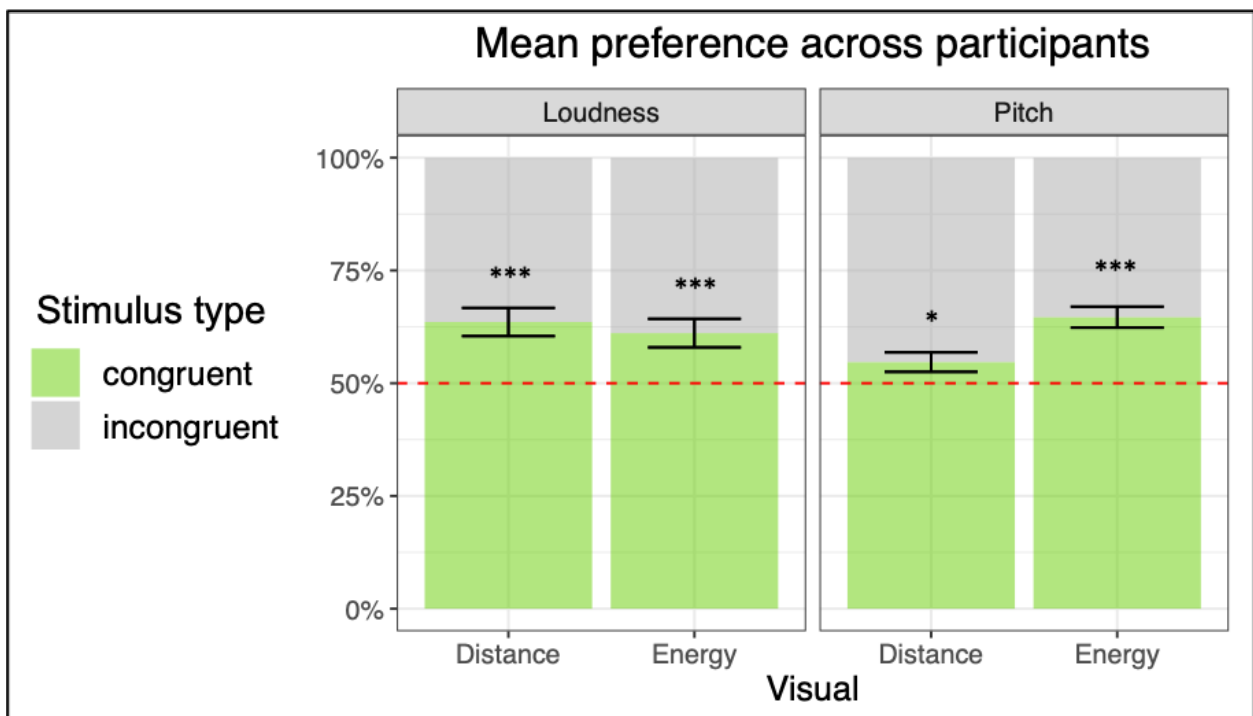


Figure 3: Results from experimental block 1 on the interpretation of pitch and dynamics as distance and energy

4 Testing coreference with pitch and dynamics

Having established the experimental reality of semantic interpretations of pitch and dynamics, we will now ask whether they can help establish reference to objects.

4.1 Experimental paradigm

The goal of this second block of the experiment was to determine whether dynamics and pitch could establish reference relations to objects. The theory tested here, referred to as the target theory hereafter, could be phrased as follows:

Target theory: Musical properties are associated with objects as soon as these objects satisfy the interpretation of these properties.

In other words, when one musical sound and two objects are present, reference is established between the musical sound and an object if that object satisfies the interpretation of the property.

4.1.1 Target stimuli

To clarify the design of the stimuli, let us derive the predictions from this target theory for [this particular stimulus](#), which can be schematically and discretely represented as follows:

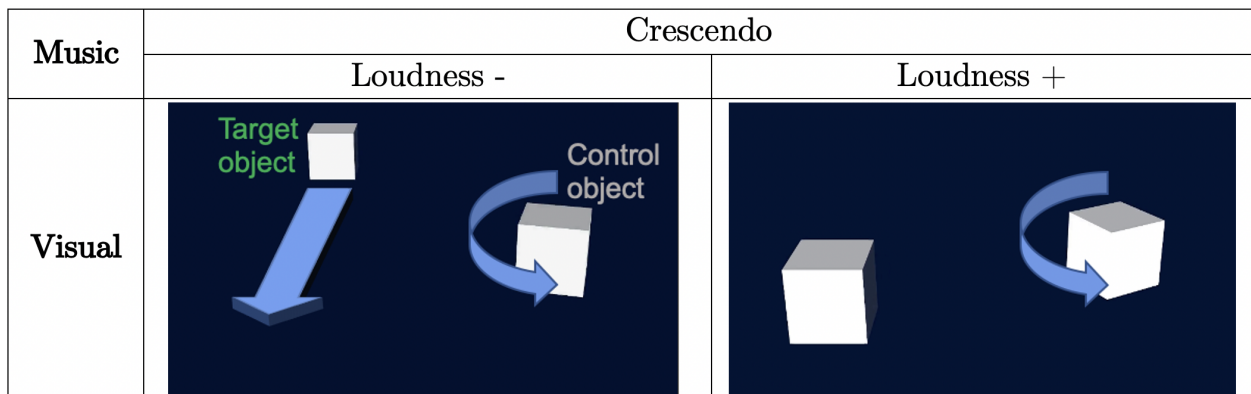


Figure 4: A discrete representation of an audiovisual stimulus

This audiovisual stimulus involves a sound with a crescendo (increasing loudness), constant pitch and constant timbre. Visually, it shows two white cubes that are identical in all respects, except for their position and movement. One, on the left, is moving closer to the observer at

constant speed and thus satisfies the crescendo (referred to as the target object hereafter), and the other one, on the right, is spinning on itself (referred to as the control object). The reason for having both objects move was to make sure that participants do not merely associate the sound with the moving object, regardless of whether that object satisfies the musical property.

Participants were asked to pick the object they thought the sound referred to. The experimental interface looked like the screenshot displayed in Figure 5. In our example, the prediction was that the cube on the left, moving closer and matching the crescendo, would be preferably selected. All stimuli used for this second block of the experiment, accessible in Tables 2 and 3, were designed in the same way as the example discussed above. The stimuli testing for loudness involved either (i) a crescendo paired with a target object approaching or gaining energy, or (ii) a decrescendo paired with a target object either moving away or losing energy. The stimuli testing for pitch involved either (i) an ascending scale in major or minor mode paired with a target object gaining energy, or (ii) a descending scale in major or minor mode paired with a target object losing energy.¹⁵ Just as in the first block, energy gain/loss was visually indicated through a shaking movement of the target object with a linearly increasing/decreasing amplitude. The control object was always spinning. Just as for the first block, each stimulus had a cube version and a cylinder version in order to make sure that the evidenced effects, if any, were not specific to a specific 3D object and to introduce variability in the stimuli. The position of the target object was randomized.

4.1.2 Control stimuli

In the event that, as predicted, a preference for the target object (the one satisfying the musical property, here the cube on the left) was found, it would still be possible that this preference is found merely due to (i) a salience effect (the cube moving closer is more salient in the visual scene than the cube spinning) or (ii) a general preference for the behavior of that object (maybe people just prefer objects that move towards them). This is the reason why two conditions were introduced. Stimuli in the **target condition** consisted of scenes

¹⁵Note that in this second block, we did not test for the ability of pitch to establish reference to the target object through its interpretation as distance, for we considered that not enough theoretical background justified predictions in this regard

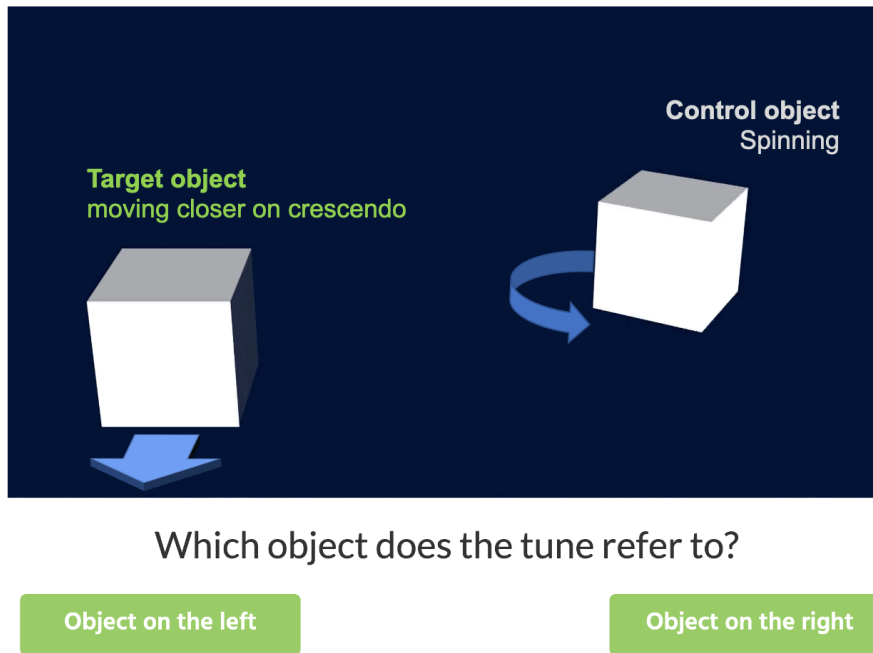


Figure 5: Experimental task - 2nd experimental block

with a musical sound changing with respect to one of the two tested properties (dynamics or pitch): the stimulus represented in Figure 4 falls under this condition. Stimuli in the **control condition** consisted of ambiguous scenes with a constant discrete musical sound instead (neither changing in dynamics, nor in pitch).¹⁶ The experimental design for this second experimental block with accessible stimuli is provided in Tables 2 and 3.

If our target theory is true, we predicted a greater preference for the target object in the target condition than in the control condition. If not, then the prediction was that the preference level for the target object should be the same in the target and in the control conditions.

4.2 Results

As predicted, the results in Figure 6 reveal a significantly higher preference for the target object in the target condition than in the control condition, for both interpretations of

¹⁶Note that the target and control *conditions* are different from the target and control *objects*: the difference between the conditions lies in the *sound* (changing/constant), while the difference in the objects lies in their movement (energy or distance change/spinning). All stimuli, in both the target and control conditions, involved a target and a control object.

		Target object			
		Moving closer	Moving away	Gaining energy	Losing energy
Sound	Crescendo	[-]		[-]	
	Decrescendo		[-]		[-]
	Constant	[-]	[-]	[-]	[-]

	Target condition
	Control condition

Table 2: Distribution of the stimuli across conditions in experimental block 2 - DYNAMICS
(All stimuli can be accessed by clicking on the corresponding cell.)

		Target object	
		Gaining energy	Losing energy
Sound	Ascending pitch (major)	[-]	
	Ascending pitch (minor)	[-]	\
	Descending pitch (major)		[-]
	Descending pitch (minor)	\	[-]
	Constant	[-]	[-]

	Target condition
	Control condition

Table 3: Distribution of the stimuli across conditions in experimental block 2 - PITCH
(All stimuli can be accessed by clicking on the corresponding cell.)

dynamics (as distance and as energy).¹⁷ Although participants tended to prefer the target object even when a constant sound was heard, they picked the target object significantly more when it satisfied the musical sound.

¹⁷For loudness, a linear mixed-effect model revealed a significant effect of stimulus type (control vs target) ($\chi^2 = 73.928, p < 2.2e - 16$), and similarly for pitch ($\chi^2 = 13.658, p = 0.0002193$). As mentioned in the scripts, the simplification of the models made due to the preregistered models failing did not allow for testing the significance of the interaction between stimulus type (target/control) and visual interpretation (distance/energy) in the case of dynamics. Qualitatively, at least, this interaction is not visible.

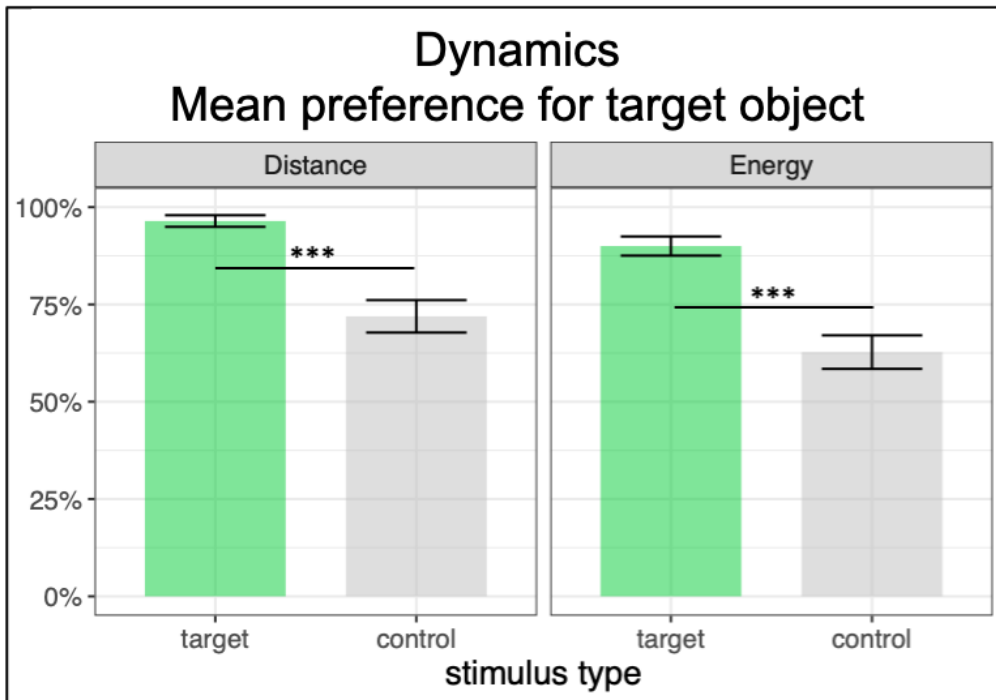


Figure 6: Results from experimental block 2 - dynamics

The same pattern was found for pitch as seen in Figure 7: we found a significantly higher preference for the target object in the target condition, compared to the control condition.

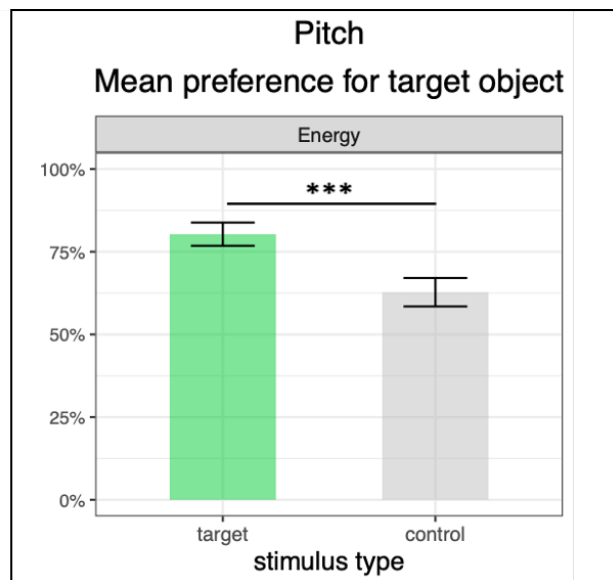


Figure 7: Results from experimental block 2 - pitch

Together, these results confirm our target theory that both pitch and dynamics establish reference relations to objects. In the target condition, one object only satisfied the semantics

of the sound, and that object was systematically judged as being the one the sound was about. Importantly, this pattern can neither be explained by a mere salience effect of the object, nor by the fact that it was the only one moving in the scene. Hence, we can conclude that the visual representation of the target object and the musical sound are coreferential: they both refer to a same object.

5 Testing coreference with timbre

This final section investigates how timbre behaves with regards to reference to objects. We will here argue that (i) timbre encodes reference to objects just as pitch and loudness, and that (ii) timbre makes it possible to show that reference relations hold between musical properties and *objects* (and not only between musical properties and *scenes* involving those objects, a distinction that the paradigm in Section 4 was unable to make).

5.1 Motivation

Although the results in section 4 confirmed our target theory, they seem to be unable to reject an alternative theory which would yield the same predictions. We will first present this alternative theory, and then explain how manipulations of timbre can be used to reject it.

Alternative theory

Musical properties are associated with *scenes* as soon as some objects within these scenes satisfy the interpretation of these properties.

This alternative theory predicts that as long as *some* part of the visual scene satisfies the sound property, the scene is considered felicitous. This does not entail a proper binding of musical properties and objects, but merely an association of musical properties to a weak notion of objects defined as parts of scenes. It is therefore impossible to assert from the previous results that pitch and loudness establish reference to objects strictly speaking, but only that they refer to scenes involving objects.

From an experimental standpoint, it was however not possible to reject the alternative theory using the paradigm described in 4 because the task was specifically to answer a question about reference to an object ('Which object does the tune refer to?'). But if our

alternative theory is true, and music rather refers to situations than objects, then it is possible that participants simply responded by disregarding whatever in the scene did not match the music (the control object), and hence reconstructed a notion of object from the scene. If that is the case, then despite the contrasts found in 4, the results do not show that musical properties referred to objects.

What was needed then was to design stimuli so that no part of the scene could be disregarded. One way of doing so is to have both objects satisfy the musical property, so that neither can be ignored. But if both objects satisfy the musical property, then both can be denoted by that property (by definition), and the reference relation is lost: it would be impossible to know which of the two objects is being referred to.¹⁸ Since it is not possible to test reference to objects rather than to scenes with just one property and two objects satisfying it, we reasoned that introducing a *second* property makes this possible, for it allows both objects to satisfy a first property, and only one to satisfy a second property so as to preserve reference.

This is where timbre comes into play. Although theoretically, any combination of two musical properties should allow for different predictions of the two theories (such as dynamics and pitch, for instance), timbre seemed like a particularly good candidate second property, for it has been argued to be responsible for object identification (i.e. interpreted as the nature of an object (Schlenker, 2022)). Our target theory applied to timbre therefore states that timbre is associated with an object as soon as this object satisfies the interpretation of timbre. Given that the interpretation of timbre is precisely the identity of the object it is associated with, satisfying the interpretation amounts to having this association consistent throughout the scene, and the prediction is that *Timbre is associated with an object as soon as this association is consistent*. By contrast, the alternative theory states that timbre is associated with a *scene* as soon as some object within that scene satisfies the interpretation of timbre, which by the same reasoning makes the prediction that *Timbre is associated with a scene as soon as some object is consistently present in that scene*.

¹⁸The same problem arises if we think about both objects satisfying the musical property at different times.

Note that the main difference in the predictions of the two theories is that the target theory requires that the timbre/object association be consistent, while the alternative theory does not; it only requires that the scene consistently contains some object. We explain how we designed our stimuli based on this difference in the following section.

5.2 Experimental design

Each stimulus was designed as a scene involving two objects, each of which was assigned a given timbre during a pairing phase. The reason for this pairing phase was that to create differences in the timbre/object association consistency, it was first needed to create that association. It was established through the repetition of co-occurrences of forward movements satisfying a crescendo and paired with corresponding timbres. Each object was either paired with a cello timbre or an organ timbre (available in Garageband[©] instruments library). This choice was made so as to guarantee a sharp contrast and a clear identification of each timbre. The resulting scene is discretely represented in Figure 8.






Audiovisual events		E_1	E_2	E_3	E_4	E_5
Visual						
Sound Crescendo	match	Timbre 1	Timbre 2	Timbre 1	Timbre 2	Timbre 1
	mismatch	Timbre 1	Timbre 2	Timbre 1	Timbre 2	#Timbre 2
PAIRING PHASE						FINAL PHASE

Figure 8: Design and stimuli - 3rd experimental block

At any point in time during the pairing phase (in events E_1 to E_4), one of the two objects satisfies the crescendo. In the final phase (event E_5), the last object to move forward always satisfies the crescendo as well. However, this object only satisfies the condition on timbre established during the pairing phase (consistency) *in the ‘match’ condition* (an example is accessible [here](#)). In the ‘mismatch’ condition, the last object to move was paired with a sound whose timbre had previously been associated with another object (inconsistency) (the corresponding stimulus violating the timbre/object association is accessible [here](#)).

If the alternative theory is true, i.e. if musical properties are associated with the scene as soon as some object in that scene satisfies these properties, then we predict that both kinds of stimulus should be equally good and that no preference should be found for either one: in both conditions, an object satisfies the final crescendo. The alternative theory does not make any prediction as to *which* of the two objects should satisfy this final crescendo because it does not require that timbre/object associations be consistent.

By contrast, if the target theory is true and the alternative theory is false, then stimuli in the match condition should be preferred because the object that satisfies the final crescendo also has to satisfy the condition on timbre. In other words, if timbre combined with dynamics does establish reference to objects and not merely to scenes, then the stimuli with coreference holding between both properties should be preferred.

Participants were presented with pairs of videos, one from the match condition, and the corresponding one from the mismatch condition. Audiovisual stimuli were identical in every respect except for the last event in the scene, which was paired with different timbres as shown in 8. Participants had to click on the video they preferred.¹⁹ The appearance of the experimental interface was very much similar to that shown in Figure 2. The position of each audiovisual stimulus was fully randomized; as was the position of the object moving first within the videos themselves. The stimuli were designed so that half of them involved the repetition of a same timbre in events E_4 and E_5 and half of them did not, to avoid biases from the repetition of a sound with a same timbre directly associated with another object. The order of stimuli was fully randomized as well.

5.3 Results

As predicted, the results in Figure 9 show a statistically significant preference for the coreferential stimuli over the non-coreferential ones.²⁰ In other words, participants tended to prefer the interpretation under which the object/timbre pair remained constant over

¹⁹Note that the task had to change to test for reference to objects indirectly, because as mentioned earlier, asking participants about which object refers to already presupposes reference to object and does not allow for a dissociation in predictions between a music/scene and a music/object pairing.

²⁰Just as for the first experimental block, we just performed a one-sample t-t-test to check whether coreferential stimuli were preferred significantly more than 50% of the time, which it did: $t = 4.1087, df = 49, p - value = 0.0001508$

the interpretation under which that pairing changed. These results allow us to reject the alternative theory: it is not the case that timbre is merely associated with scenes containing objects that satisfy them; timbre *and* loudness are in this case associated with the objects themselves. Therefore, the associations from music to objects evidenced in this paper are more than just correlations to properties of scenes: they are reference relations to *objects*. And because the musical sound referred to the same object as the object represented by the visual object in the scene, we can conclude that both representations (musical and visual) were coreferential.

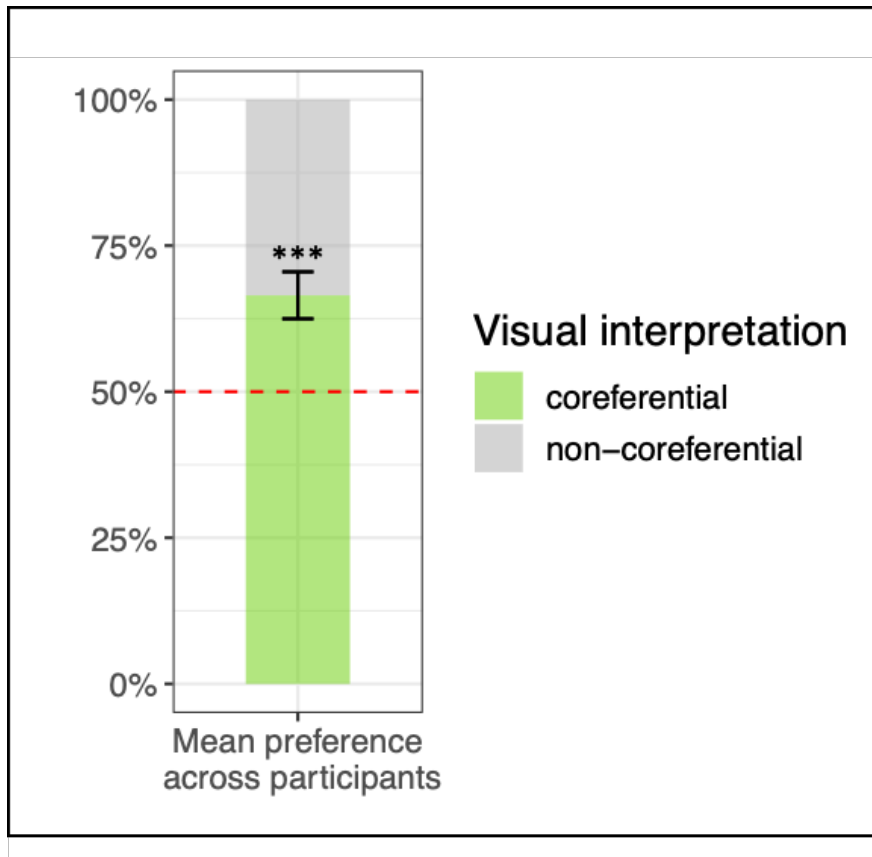


Figure 9: Results from experimental block 3
Preference for coreferential over non-coreferential interpretation

6 Conclusion

We can draw three conclusions from the previous results. First, it is indeed possible for pitch and dynamics to be interpreted as energy levels and/or distance to the observer. We now have experimental evidence that the three following rules hold:

1. The higher the pitch, the higher the level of energy of the object.
2. The louder the dynamics, the higher the level of energy of the object.
3. The louder the dynamics, the closer the object to the observer.

We also gathered significant data in favor of the following rule, which will yet have to be further investigated, both theoretically and experimentally:

4. The higher the pitch, the closer the object to the observer.²¹

Second, we confirmed that it is possible for pitch and dynamics to determine reference to objects: both properties can help decide which object is being referred to by the sound.²²

Third, we confirmed that timbre is able to establish reference as well when combined with other properties such as dynamics. Timbre also helped establish that reference relations hold between musical properties and *objects*, and not only between musical properties and *scenes* involving these objects.

In our experimental paradigm, we worked with minimal audiovisual stimuli that involved very simplified visual scenes and very minimal musical stimuli. This choice was deliberate as it allowed for greater simplicity in the stimuli and for better control. However, this begs the question of (i) whether similar effects are found in music without visual animations, and

²¹Although this finding did not have backing by theory, it is possible to speculate on a few reasons why we found a pitch height/distance mapping. The experimental results on associations between pitch height and depth or distance has mainly been inconsistent in the literature. While, contrary to our results, Eitan and Granot (2006) found that rises were associated with *increasing distance* (i.e. objects moving away), Eitan and Timmers (2010) did not find any correlation between high-pitch music and judgments of distance in listeners. One must however note that Eitan and Timmers (2010) did not test pitch changes but rather associations between pitch *range* and judgments of distance, while in our case, a crucial property of the sounds we tested was that they were changing in pitch. Our results are however consistent with other theories (discussed in Eitan (2013)) on the pitch/distance association modeled after the Doppler effect, which does predict that a rise in pitch should be associated with an increase in distance, which happens naturally when a sound source gets closer to a listener. They are also consistent with findings from Ghazanfar and Maier (2009), which established increased attention in Rhesus Monkeys when exposed to sounds rising in pitch, suggesting that a rise in pitch might be an important cue for detecting objects approaching in this species. Similar mechanisms seem to indirectly exist in humans: for instance, Ilie and Thompson (2006) show that low pitches are judged as more pleasant and trigger less arousal in listeners compared to high pitches. It might be that watching an object approach triggers some kind of alert system associated with higher energy arousal: the higher the pitch, the more threatening the object. From an evolutionary perspective, it might thus be that pitch is interpreted as a cue to object detection. However, one must note that the pitch/distance association was the weaker we found in the results. Although significant, this calls for caution regarding their interpretation. Further work will have to develop more fine-grained theories of pitch height/depth mappings to confirm whether our findings can be replicated, or whether this association might just be a byproduct of a more fundamental law instantiated in another space than the 3D space.

²²Although we used an audiovisual paradigm throughout this experiment, one must note the main motivation for doing so was to limit noise: even though listeners do get spontaneous mental representations from music, music semantics is so abstract that many such representations are licensed given one musical passage or piece. This fundamental property of music semantics thus makes it hard to investigate subtle effects in which the number of possible associations with objects should be fixed. It is however likely that the role the visual modality played here can be filled by other modalities, or else by the auditory modality and other musical information just as much (for instance providing context, or limiting the number of possible denotations).

(ii) whether our results generalize to cases of real music, which is far more complex than the minimal stimuli used in our experiment.

(i) Our use of visual animations paired with musical stimuli was motivated by the fact that music often licenses many mental scenes: to a same musical excerpt correspond many situations involving many different objects. Formal accounts of musical meaning have claimed that there are many situations that a same musical excerpt *is true of* (Schlenker, 2017, 2022). This results in musical meaning being very abstract: it is of course possible that different listeners spontaneously represent music in different ways, i.e. with different corresponding situations. But a central claim has been that all these situations do share some of their properties, and using visual animations is a way of making these situations *less* abstract: it is a methodological reason that motivates work with audiovisual scenes. Let us take an example from the stimuli used in this experiment: a crescendo. We have seen that a crescendo can be interpreted as an object approaching. We have already seen from the data that the results are insensitive to whether the object was a cube or a cylinder. Arguably, they would remain the same if it were any other object.

It is however another question whether they would be preserved without any visual cue. In the case of the first block, they probably would. If the assumption that listeners interpret music as being *about* something else than music itself is true, then it is likely that listeners have the ability to represent objects even in the absence of a visual anchor. In the case of the second block, the contribution of the visual scene was to force the representation of a scene where two objects were present instead of one. But if participants could consistently decide which of the two objects was being referred to, then it is highly likely that the same results would have been found whether three, four or more objects were present, as long as one object only satisfied the musical property. And if this happens with any number of objects, then it follows that regardless of the number of objects one posits, the reference to one object is preserved as long as that object matches the musical properties. In the third block, the contribution of the visual scene was to force an interpretation under which two objects were present. The results from this block showed that the stimuli where the timbre/object relation was consistent were preferred, which suggests that timbre might well

be interpreted as the identity of an object. This is consistent with the idea that there are as many objects represented as there are timbres, which is in line with a principle stated in Schlenker (2022) that ‘When two musical events have different timbre, they tend to be not coreferential.’ It is therefore plausible that in the absence of any visual information, listeners would still represent two objects when two different timbres are heard. However, the point about the reference to objects could not have been made without a visual contribution because music alone is more permissive than audiovisual scenes. Using audiovisual scenes can thus help reveal mechanisms that (probably) exist in music alone (such as reference to objects rather than scenes), but that are either hard or impossible to test in the absence of any interaction with another representational system.

(ii) Under the assumption that our results hold for music alone, a final question as to whether they generalize to real music remains. It is true that we worked with very simplified music which hardly approximates the level of complexity of real music. However, we argue that if such effects can be found in minimal stimuli, then they should be at least partially preserved in more sophisticated music that makes use of similar patterns. Crescendos and rises in pitch are very frequent in music, and although they might interact with other properties that this experiment did not explore (such as harmonic structure, tempo, subtle articulations or interpretative adjustments), there is no reason to believe that the semantic effects we gave evidence for here should disappear. However, if one wanted to get closer to real music, one way to do it is to reconstruct a musical passage through similar rules as the ones we tested in Section 3: each combination of properties should come with its own set of rules. It is possible that in real music, many more than two objects are often represented; and it is also possible that some of the tendencies and rules we discussed here do not hold in certain cases. Although the *scenes* that real music can be true of certainly have a greater level of complexity, this is arguably independent from the fact that musical properties are mapped onto objects and not merely onto scenes. This fundamental property of music to refer to objects should therefore not be lost as music gets more complex; but further work should examine whether this claim really holds in real music.

Acknowledgments

This research received funding from the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme (grant agreement No 788077, Orisem, PI: Schlenker). Research was conducted at Institut d'Etudes Cognitives, Ecole Normale Supérieure - PSL University. Institut d'Etudes Cognitives is supported by grant FrontCog ANR-17-EURE-0017.

Ethics

This research is part of the project review and approved by the ethics committee of Inserm, the Institutional Review Board (IRB00003888, IORG0003254, FWA00005831) of the French Institute of medical research and Health under Opinion number 20-733.

Bibliography

- Abusch, D. (2013), ‘Applying Discourse Semantics and Pragmatics to Co-reference in Picture Sequences’, *Proceedings of Sinn und Bedeutung* **17**, 9–25.
- Abusch, D. (2017), The formal semantics of free perception in pictorial narratives., *in* ‘Proceedings of the 21st Amsterdam Colloquium’, a. cremers, t. van gessel, & f. roelofsen edn, pp. 85–96.
- Abusch, D. (2020), Possible-Worlds Semantics for Pictures, *in* ‘The Wiley Blackwell Companion to Semantics’, John Wiley & Sons, Ltd, pp. 1–31. _eprint: <https://onlinelibrary.wiley.com/doi/pdf/10.1002/9781118788516.sem003>.
- Bachorowski, J. A. (2008), Vocal expressions of emotion., *in* ‘Handbook of emotions, 3’, pp. 196–210.
- Blumstein, D. T., Bryant, G. A. and Kaye, P. (2012), ‘The sound of arousal in music is context-dependent’, *Biology Letters* **8**(5), 744–747.
- Clarke, E. (2001), ‘Meaning and the Specification of Motion in Music’, *Musicae Scientiae* **5**(2), 213–234.
- Cross, I. and Woodruff, G. E. (2009), Music as a communicative medium, *in* R. Botha and C. Knight, eds, ‘The Prehistory of Language’, Oxford University Press, pp. 77–98.
- Eitan, Z. (2013), How pitch and loudness shape musical space and motion, *in* ‘The psychology of music in multimedia’, Oxford University Press, New York, NY, US, pp. 165–191.
- Eitan, Z. and Granot, R. Y. (2004), Musical Parameters and Spatio-Kinetic Imagery, *in* ‘ICMPC8 : proceedings of the 8th international conference on music perception & cognition: August 3-7, 2004 : Evanston, Illinois, USA’, Causal Productions, pp. 57–63.
- Eitan, Z. and Granot, R. Y. (2006), ‘How Music Moves’, *Music Perception* **23**(3), 221–248.
- Eitan, Z. and Timmers, R. (2010), ‘Beethoven’s last piano sonata and those who follow crocodiles: Cross-domain mappings of auditory pitch in a musical context’, *Cognition* **114**, 405–422. Place: Netherlands Publisher: Elsevier Science.
- Gabrielsson, A. and Lindström, E. (2010), The role of structure in the musical expression of emotions, *in* ‘Handbook of music and emotion: Theory, research, applications’, Series in affective science, Oxford University Press, New York, NY, US, pp. 367–400.
- Ghazanfar, A. A. and Maier, J. X. (2009), ‘Rhesus monkeys (*Macaca mulatta*) hear rising frequency sounds as looming.’, *Behavioral Neuroscience* **123**(4), 822–827.
- Godøy, R. I. and Leman, M., eds (2010), *Musical gestures: sound, movement, and meaning*, Routledge, New York. OCLC: ocn298781501.
- Ilie, G. and Thompson, W. F. (2006), ‘A Comparison of Acoustic Cues in Music and Speech for Three Dimensions of Affect’, *Music Perception* **23**, 319–329. Place: US Publisher: University of California Press.
- Ionin, T. and Zyzik, E. (2014), ‘Judgment and Interpretation Tasks in Second Language Research’, *Annual Review of Applied Linguistics* **34**, 37–64. Publisher: Cambridge University Press.
- Juslin, P. N. and Laukka, P. (2003), ‘Communication of emotions in vocal expression and music performance: Different channels, same code?’, *Psychological Bulletin* **129**(5), 770–814.
- Karttunen, L. (1969), Pronouns and Variables., R. I. Binnick, A. Davison, G. M. Green, and J. L. Morgan., Chicago: Department of Linguistics, University of Chicago. Reprinted in J.

- D. McCawley, ed., *Syntax and Semantics 7: Notes from the Linguistic Underground*. New York: Academic Press, 1976.
- Koelsch, S. (2012), *Brain and music*, Wiley-Blackwell, Chichester, West Sussex ; Hoboken, NJ. OCLC: ocn767563922.
- Larson, S. (2012), *Musical forces: motion, metaphor, and meaning in music*, Musical meaning & interpretation, Indiana University Press, Bloomington. OCLC: ocn707212791.
- Montague, R. (1973), The Proper Treatment of Quantification in Ordinary English., in ‘Approaches to Natural Language.’, P. Suppes, J. Moravcsik, and J. Hintikka, Reidel: Dordrecht.
- Olsen, K. N. (n.d.), ‘Intensity dynamics and loudness change: a review of methods and perceptual processes’, *Acoustics Australia* .
- Patel-Grosz, P., Grosz, P. G., Kelkar, T. and Jensenius, A. R. (2018), ‘Coreference and disjoint reference in the semantics of narrative dance’, *Proceedings of Sinn und Bedeutung* **22**(2), 199–216. Number: 2.
- Schlenker, P. (2017), ‘Outline of Music Semantics’, *Music Perception* **35**(1), 3–37.
- Schlenker, P. (2022), ‘Musical meaning within Super Semantics’, *Linguistics and Philosophy* **45**(4), 795–872.
- Schlenker, P., Bonnet, M., Lamberton, J., Lamberton, J., Chemla, E., Santoro, M. and Geraci, C. (2023), ‘Iconic Syntax: Sign Language Classifier Predicates and Gesture Sequences’. LingBuzz Published In: To appear in *Linguistics & Philosophy*.
- Seyfarth, R. M. and Cheney, D. L. (2003), ‘Meaning and Emotion in Animal Vocalizations’, *Annals of the New York Academy of Sciences* **1000**(1), 32–55. _eprint: <https://onlinelibrary.wiley.com/doi/pdf/10.1196/annals.1280.004>.
- Sievers, B., Lee, C., Haslett, W. and Wheatley, T. (2019), ‘A multi-sensory code for emotional arousal’, *Proceedings of the Royal Society B: Biological Sciences* **286**(1906), 20190513.
- Sievers, B., Polansky, L., Casey, M. and Wheatley, T. (2013), ‘Music and movement share a dynamic structure that supports universal expressions of emotion’, *Proceedings of the National Academy of Sciences* **110**(1), 70–75.

Chapter 4

Meaning from music alone

Purpose

In the previous chapter, we saw that investigating the interaction between musical sounds and visual perception reveals interesting properties of music semantics. In this chapter, we move away from paradigms using music in combination with language or visual perception, and explore a pure version of musical meaning through a case study of music evoking a particular instance of body motion: walking. Our findings show that investigating such specific case studies help reveal some further properties of music semantics.

This chapter contains two articles. The first one, which was published in the Proceedings of the 22nd Amsterdam Colloquium (2019) [reference provided in the paper], explores theoretical implications of walk-denoting music. The second article presents the detailed results from an experiment aiming at testing the predictions from the theoretical paper. Most of the theoretical content from the first paper is explained in the second paper as well.

WALK-DENOTING MUSIC

REFINING MUSIC SEMANTICS *

Léo Migotti
Institut Jean Nicod
ENS-EHESS-CNRS

Léo Zaradzki
Laboratoire de Linguistique Formelle
Université Paris Diderot

ABSTRACT

Music has recently been argued to have a referential semantics (Schlenker, 2017, 2019), *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.

*This paper was published in the Proceedings of the 22nd Amsterdam Colloquium:
Migotti, L. Zaradzki, L. (2019). *Walk-denoting music: refining music semantics*. In Julian J. Schlöder, J.J., McHugh, D. Roelofsen, F. (eds), Proceedings of the 22nd Amsterdam Colloquium, pp. 593-602. [\[link\]](#)

Contents

1	Introduction	116
2	A formal model for music semantics	116
3	Walk-denoting music and walking-situations	118
4	Theoretical refinements	121
4.1	Hierarchical structures	121
4.2	Do all notes have the same semantic status?	124
4.3	The meaning of uninterpreted notes	127
4.4	Is every event musically represented?	128
5	Conclusion	128
	Bibliography	130

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, 2017, 2019), 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, 2019). 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.²

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 (2017, 2019) 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

²We 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.

meaning. The following section draws from Schlenker (2017, 2019) 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, 2017, 2019). 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*³.

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

³Their 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.

to the set of situations a musical snippet can evoke. Formally, we define, just as Schlenker, a musical snippet as an n -tuple $M = (m_1, \dots, m_n)$, and a possible situation S as an n -tuple $S = (e_1, \dots, e_n)$, where (m_1, \dots, 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 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 1⁴.

⁴audio file: <https://www.youtube.com/watch?v=BWbqZ1BiRzI&feature=youtu.be>
All scores are directly clickable to access the audio file.

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 2⁵.

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 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.



Figure 1: Prototypical walk-denoting music

⁵audio 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 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 conditions would trigger an inferential preference for the prototypical snippet, that satisfies all conditions⁶.

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

⁶In 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>.

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 4⁷. Indeed, introspective judgments given by our informants as well as our own 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 it was affected in a very subtle way, that did not correspond 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 situation⁸. 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

⁷audio file: <https://www.youtube.com/watch?v=gR-wHm4nFYk&feature=youtu.be>

⁸While 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.

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 listeners’ common intuitions, for the opening theme of Mozart’s 40th 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, \dots, e_k and a are events of S associated in M with musical events m_1, \dots, m_k and b , and if there is a group to which e_1, \dots, 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



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

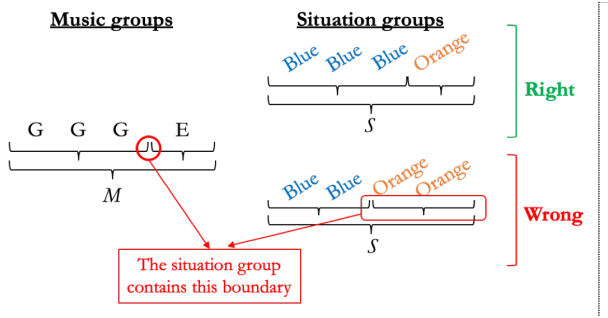


Figure 6: Example of grouping mismatch.

belongs. As an illustration of this phenomenon, let us take the opening motif of Beethoven’s 5th Symphony. This consists in four notes: G G G E. According to Lerdahl and Jackendoff’s rules⁹ they are grouped as $[[G, G, G], E]$. In the corresponding section of *Fantasia 2000*¹⁰ 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 level¹¹. 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'

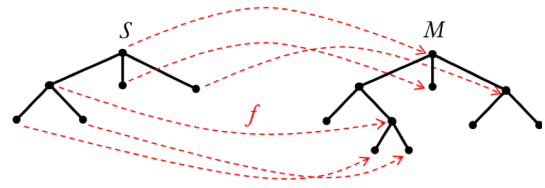


Figure 7: Example of tree embedding

⁹and also to general Gestalt principles

¹⁰an animated film where image is intended to be a denotation of the music; see: <https://www.youtube.com/watch?v=nMn1xYkZKaU>

¹¹We 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.

represent the sets of vertices and E and E' represent the sets of edges¹², a necessary condition for $M \models S$ is the existence of an injective root-preserving function $f : V' \rightarrow V$ 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** (strong¹³ 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 situation¹⁴.

¹²Since the tree is directed, edges are ordered pairs of vertices.

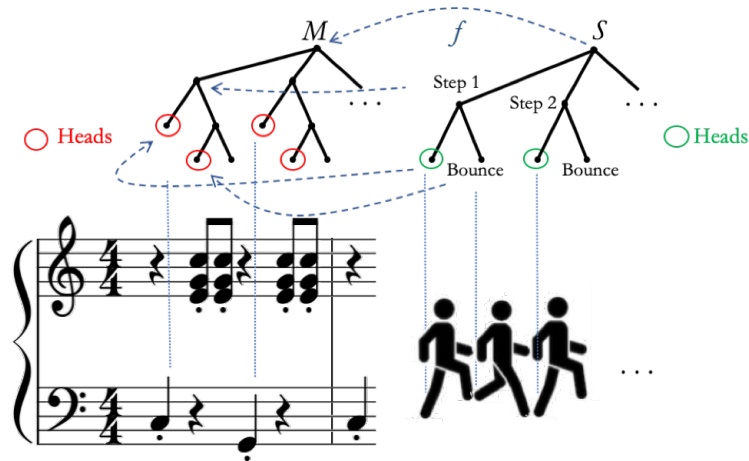


Figure 8: An illustration of Theory B on the example from Figure 4.

- **Theory B** (weak¹³ 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 situation¹⁴.

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

¹³The names *weak* and *strong* theories reflect the fact that Theory A requires more constraints.

¹⁴This last idea is related to what Schlenker has suggested in Section 8.5 of Schlenker (2019). 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.

so on. Since the number seems to be unbounded, we rather replace the notion of heads by that of *weight*¹⁵. 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 $p : V \rightarrow \mathbb{R}_+^*$ (resp. $p' : V' \rightarrow \mathbb{R}_+^*$). 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 $\forall x, y \in V$, $p(x)p(y) \Rightarrow p'(g(x))p'(g(y))$. 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-command¹⁶. Formally, our localised condition becomes: $\forall y \in V$, $(\exists x \in \text{Im}f \text{ s.t. } y \text{ c-commands } x \text{ and } p(y)p(x)) \Rightarrow y \in \text{Im}f$. We also add the weight-preservation condition as in the other theory.

An alternative notion to weight would be that of reduction¹⁷. 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

¹⁵though the same intuition is behind

¹⁶Abbreviation 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.

¹⁷At 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.

— as it is the case of the structures that Lerdahl and Jackendoff (1983) deal with¹⁸ — 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 speed¹⁹. 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 variant²⁰, 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

¹⁸The 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.

¹⁹Typically, 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.

²⁰Thanks to Philippe Schlenker for suggesting this variant.

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.²¹ 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 tree²² 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 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.

Theorem 1 *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 $f : V' \rightarrow V$ such that:*

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

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

²¹Think, for example, of the whole subtlety in the gestures of a character, or of all the particle interactions that take place everywhere.

²²Just like there are notions of reduction for music, we can posit that the same things exist with situations, as suggested in Jackendoff (2009).

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.

Bibliography

- Harasim, D., Rohrmeier, M. and O'Donnell, T. J. (2018), 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', pp. 152–159.
- Jackendoff, R. (2009), 'Parallels and nonparallels between language and music', **26**(3), 195–204.
- Lerdahl, F. and Jackendoff, R. (1983), *A generative theory of tonal music*, The MIT Press, Cambridge, MA.
- Schlenker, P. (2017), 'Outline of music semantics', **35**(1), 3–37.
- Schlenker, P. (2019), 'Prolegomena to music semantics', **10**(1), 35–111.
- Zacks, J., Tversky, B. and Iyer, G. (2001), 'Perceiving, remembering, and communicating structure in events', *Journal of experimental psychology. General* **130**, 29–58.

HOW MUSIC EVOKES WALKING

A CASE STUDY IN MUSIC SEMANTICS *

Léo Migotti
Institut Jean Nicod
ENS-EHESS-CNRS

Léo Zaradzki
Laboratoire de Linguistique Formelle
Université Paris Diderot

ABSTRACT

Music has recently been claimed to have a semantics, i.e. it is able to refer to a non-musical reality in systematic ways. In this framework, different musical properties convey different kinds of information about abstract objects or situations through semantic rules that map a given musical form onto a set of possible situations. Because music semantics is very abstract (many mental representations are often licensed by a same musical excerpt), we here chose to work with a concrete situation that music can denote: someone walking. By manipulating fine-grained musical properties of a reference stimulus, we tried to identify what it takes for a music to evoke a walk in listeners. We built a preference task in which participants had to rate which of two excerpts was better at representing a walk. Our results show that at least 5 properties are involved in making a certain music evoke a walk: while some structural constraints map onto the structure of a walking event (steadiness, chord alternation, binarity), some more specifically musical properties (consonance and tonal chord proximity) seem to be associated with physical stability. Our results provide further enrichments and updates on a possible model for music semantics, in particular on (i) the type of rules involved and (ii) the structural level at which music is interpreted.

Keywords Music semantics | Music and movement | Walking | Syntax/semantics interface

* *Contributions*: LM and LZ designed the experiment. LZ composed most stimuli which LM generated and implemented in the online experiment. LM managed online data collection and statistical analyses. LM wrote the paper, taking some inspiration from LZ's PhD dissertation on the same topic.

Contents

1	Introduction	138
2	General paradigm	139
2.1	Target stimulus	139
2.2	Task	140
3	Some musical stimuli evoke walking better than others	142
3.1	Stimuli	142
3.2	Results	142
4	Identifying musical properties responsible for the music/walk association	144
4.1	Candidate properties	144
4.2	Degraded stimuli	145
4.3	Results	147
5	Enriching the model of music semantics	148
5.1	Ordering preservation-based model	148
5.2	Wrong predictions from RPOs	150
5.3	Stimuli	151
5.4	Results	152
6	Are all musical events semantically interpreted?	154
6.1	Motivation	154
6.2	Stimulus	155
6.3	Results	157
7	General discussion and conclusions	158
A	Predictions from stimuli	162
B	Statistics	169

1 Introduction

Although it is common to assume that music triggers many cognitive effects such as emotions (Meyer, 1956; Gabrielsson and Lindström, 2010; Koelsch, 2012; Sievers et al., 2019) or expectations (Huron, 2006), the idea that it can also evoke non-musical representations in systematic ways has received less attention, despite both formal and informal accounts on how listeners spontaneously make associations between music and imagined scenes or situations (Eitan and Granot, 2006; Eitan and Timmers, 2010; Eitan, 2013). It has in particular been argued that these associations can be accounted for by semantic rules that establish connections between music and what it evokes in listeners: this amounts to arguing that music has a semantics (Schlenker, 2017). Under this view, music does not only convey information about itself or its own form, but also about an extra-musical reality: when listening to a tune, a song or a symphony, one does not only draw inferences about the actual source of the sound (the violin, the singer, the orchestra), but one is able to represent music in more abstract terms, and associate musical properties with properties of imagined scenes and objects in a productive fashion.

Here, we propose to test the idea that music has a semantics experimentally through a case-study on how music can evoke a simple movement: walking. The purpose of this paper is to understand why some musical pieces represent walks better than others, and which musical properties and semantic rules are responsible for these effects. This paper is structured as follows. Section 2 presents the main stimulus whose ability to evoke a walk was tested throughout the experiment. The next sections present our experiment in three steps.² First, Section 3 shows that the target stimulus was systematically judged as evoking a walk better than 3 other pieces. Section 4 presents 5 conditions we hypothesized were responsible for the semantic association between our musical excerpt and a walk. By destroying one condition at a time and asking for preference between degraded versions of the original stimulus and the original stimulus, we show that the latter was systematically judged as better evoking a walk. Section 5 makes a contribution to a possible model for music semantics by showing

²The data was collected in a single online experiment. Our conclusions remain unaffected by this experimental constraint; we however present the different blocks separately for the sake of clarity.

that absolute levels of consonance were interpreted in terms of absolute physical stability of the walking character. Finally, Section 6 explores the hypothesis that some musical events might not be interpreted and presents results which reject this hypothesis: without further evidence, it seems that all musical events are semantically interpreted.

2 General paradigm

2.1 Target stimulus

The main goal of the experiment was to understand why the stimulus in Figure 1 (referred to as the ‘target stimulus’) was associated with a walking event in listeners. By ‘walking event’, we mean an event involving a character or person walking. This stimulus was selected for its intuitive ability to evoke a walk; yet, we provide further justification for why this association might occur in the first place in Appendix A. The target stimulus, composed after a common pattern of a ‘walking bass’,³ consists of 4 repetitions of a same 4-beats motive: a first major C chord in its fundamental state, split into a bass C on the first beat (evoking the first step), and a higher-pitched C major chord (evoking a kind of ‘rebound’). A second major C chord in its 2nd inversion state follows, split into its bass G on the 3rd beat (evoking the second step) and the same chord on beat 4 as on beat 2. The stimulus was generated through MuseScore © [Version 3.6.2.548020600] at a given tempo of 80 bpm.⁴

³As noted in Zaradkzi (2021), this terminology does of course not determine the ability of such a pattern to denote walk per se. A walking bass only corresponds to a succession of bass notes typically played by a bass or by a piano, in which case the bass notes can be played together with higher-pitched chords off-beat. In our case, to simplify the analysis, the higher-pitched chords are placed on the 2nd and 4th beats respectively rather than off-beat, so that they can be analyzed as full musical events.

⁴80 bpm approximately correspond to a slow gait pattern, the standard reference speed level for walking being usually around 90-100bpm. Note that in this experiment, tempo was not varied at any point. Although previous research had established that people are rather good at synchronizing their walk to the beat, suggesting that tempo is a very important factor contributing to the music-to-walk mapping, we were here more interested in revealing aspects that had not been addressed in the literature on the interaction between music and gait, and in particular in the investigation of musical properties unrelated to speed.

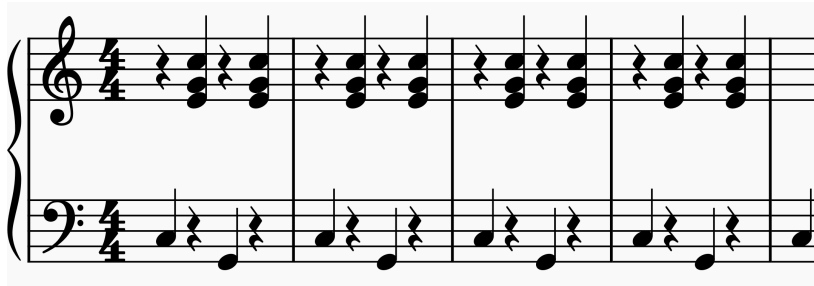


Figure 1: [Tentative target stimulus](#) ⁵

2.2 Task

50 participants took part in the online experiment, in French. They were recruited through the online platform Crowdpanel[©] and gave online consent before starting the experiment. 5 participants declared having already taken a similar study; these participants were thus removed from the data to avoid potential biases from prior exposure to a pilot study. For each trial, participants were presented two audio files that could be played as many times as needed, one of which was always the target stimulus.⁶ In each of the following sections, we present the confronting stimuli paired up with the target. Participants were instructed to imagine a character walking before playing the two stimuli, and then asked to estimate which of the two excerpts evoked a situation in which somebody was walking better, using a continuous scale ranging from the excerpt on the left to the excerpt on the right. The choice was not binary: participants could place the cursor anywhere in between the two extremities. The position of each stimulus (right or left) was systematically randomized for all pairs of stimuli. Each participant was presented 18 couples of stimuli. Details are provided in the relevant sections.

For each trial, a picture of a cartoon character walking (as shown in Figure 2) was displayed.⁷ We reasoned that this visual anchor was needed for two main reasons. First, informal pilot studies on informants suggested that having a visual substantially increased engagement with the task: it appeared that having something to look at makes the listening

⁵The stimulus is accessible by clicking on the caption of the figure directly.

⁶The stimuli are presented in the relevant sections, along with hyperlinks for the reader to listen.

⁷Our paradigm is reminiscent of previous paradigms testing for associations between music and representations of body motion, in particular Eitan and Granot (2006) where participants had to imagine a character moving through space.

experience easier, especially if the visual is relevant to the audio. Second, it was important that each participant imagined the same character walking when giving their judgment on how well the music could depict such a scene. If each participant could imagine their own walking scene freely, we would not have controlled for two major biases: (i) participants could have changed the mental scene they imagined throughout the experiment, resulting in the impossibility to compare how ‘well’ each stimulus did at triggering a representation of a character walking because different stimuli could have represented different characters, walking in different ways, with different gait patterns, etc. and (ii) even if participants did stick with such a mental scene throughout the experiment, it was possible that each participant had their own, which could have been different from that of other participants.⁸



Figure 2: A screenshot of the task (with English translations in italic)

⁸Note that in any event, we have no way of knowing whether participants did indeed represent the scene with the given character; yet we take the use of the visual to be our best attempt to fix a reference level relative to which they had to judge how convincing the music they heard was at describing the scene.

3 Some musical stimuli evoke walking better than others





The first experimental step in our procedure was to confirm introspective judgments and to make sure that the interpretation of our target stimulus as a walk was salient. We thus had participants compare the target stimulus to 3 other pieces, referred to as ‘control stimuli’ hereafter. Note that it was enough, for our purposes, that this target stimulus be significantly *more* associated with a character walking than these other stimuli. Details are provided in the next section.

3.1 Stimuli

The 3 control stimuli were selected from real music for their intuitive inability to evoke a walk mainly due to (i) rhythmic irregularities and/or slow tempo (i.e. pieces whose metrics was hardly identifiable by non-musicians, with no recurrent pattern), and (ii) their use of solo piano that matched the timber and instrumental setting of the target stimulus to keep spectral differences as minimal as possible and to avoid creating parasitic inferences from timber. As detailed in Figure 3, Control 1 was taken from the first seconds of Mendelssohn’s 1st piano concerto (2nd movement); Control 2 was taken from the first seconds of Scriabine’s Impromptu op. 12 n°1; and Control 3 was from the 1st movement of Schnittke’s concerto for piano and strings. The duration of each control roughly matched that of other stimuli, although we decided to make cuts that sounded as musically natural as possible to avoid any effect due to surprise or weirdness. The target stimulus formed 3 pairs of stimuli with each control, from which only 2 were randomly presented to each participant to reduce experimental duration and avoid redundancy.

3.2 Results

As expected, the target stimulus was significantly preferred over each control, with no significant differences in endorsement between controls: when asked to choose which excerpt

Pair	Target	Control
1		<p>Mendelssohn</p> 
2		<p>Scriabine</p> 
3		<p>Schnittke</p> 

Scores retrieved from imslp.org

Figure 3: Control stimuli
 To access the stimuli: [Control 1](#), [Control 2](#), [Control 3](#)

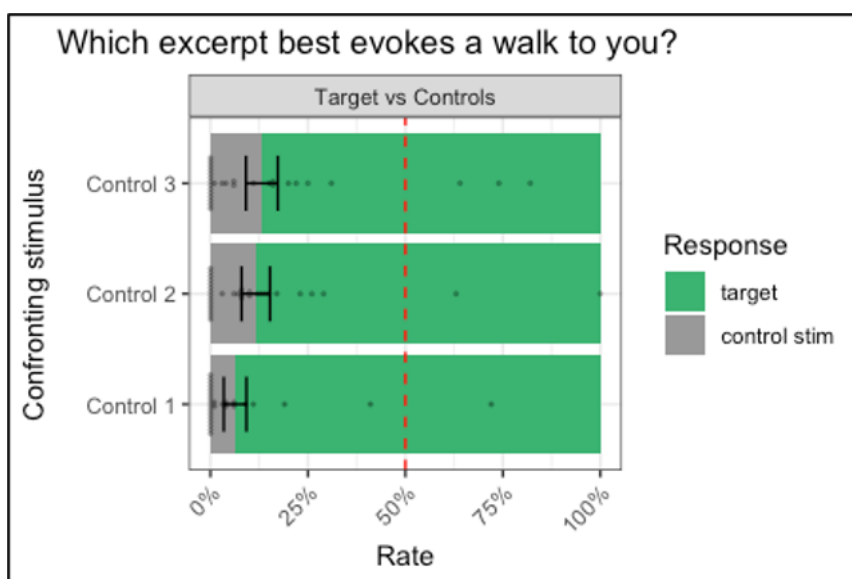


Figure 4: Preference for target over control stimuli

represented a walking event better, participants strongly tended to choose the target stimulus over any of the 3 controls.⁹

The interpretation of these results is twofold. First, they show that music has indeed the ability to refer to non-musical events.¹⁰ Second, they show that it is not a mere property of any music to be associated with any event. It is however possible that participants would not have spontaneously interpreted the target stimulus as a walk in the absence of any context. Yet, we claim that a statistically significant tendency to pair a given imagined situation (a walk) with a musical excerpt (the target) more than others is enough empirical evidence for considering the target as a good candidate for evoking a walking event.

4 Identifying musical properties responsible for the music/walk association

4.1 Candidate properties

After having confirmed that our target stimulus could be interpreted as representing a walk, we were interested in identifying the properties of the stimulus that were responsible for this effect. After careful manipulation of several musical parameters in our target stimulus, we came up with a list of 5 conditions that we hypothesized had to be met to allow any listener to associate a piece with a walking event.¹¹ These conditions are given in (1).

⁹Detailed statistical analyses are provided in B. The data and analyses script, along with pre-registered files are available at [this link](#).

¹⁰Whether the fact that music can refer to an extra-musical reality amounts to proving that music has a semantics is discussed in section 7.

¹¹Here again, the five identified conditions can also be derived theoretically: not only do we find intuitive reasons for testing these conditions through direct rewriting and testing, but these conditions also made sense given (i) a model of a walking event and (ii) a set putative semantic rules explaining the mapping from an event to music. We provide further theoretical justification in Appendix A.

(1) **Conditions on music to evoke a walking event.**

For a musical piece to be understood as referring to a walking event, it needs to be reducible to:

1. A regular
2. Alternation
3. Of 2 chords
4. That are consonant¹²
5. And tonally close to one another






We then reasoned that, if these properties were responsible for the association of the target stimulus with a walk, it was then possible to weaken or even block this association by destroying these properties. The following section details how each condition was violated through minimal rewritings of the target stimulus.

4.2 Degraded stimuli

The first manipulation was about breaking rhythmic evenness by varying the originally uniform delay between notes (DS1). The second rewriting was about breaking the alternation condition through repetition (DS2), while the third one was about breaking binarity by alternating three chords instead of two (DS3).¹³ The fourth rewriting breaks consonance by replacing major triads with dissonant clusters (DS4). The last manipulation breaks tonal proximity between the two alternating chords by increasing their distance in the tonal space (DS5). The target stimulus formed 5 pairs of stimuli with each degraded stimulus, from which only 3 were randomly presented to each participant so as to avoid redundancy throughout the experiment. Details about the rewriting procedure and the expected effect of each manipulation are provided in Figure 5.

¹²A discussion of conditions 4) and 5) and the underlying semantic model is given in Appendix B.

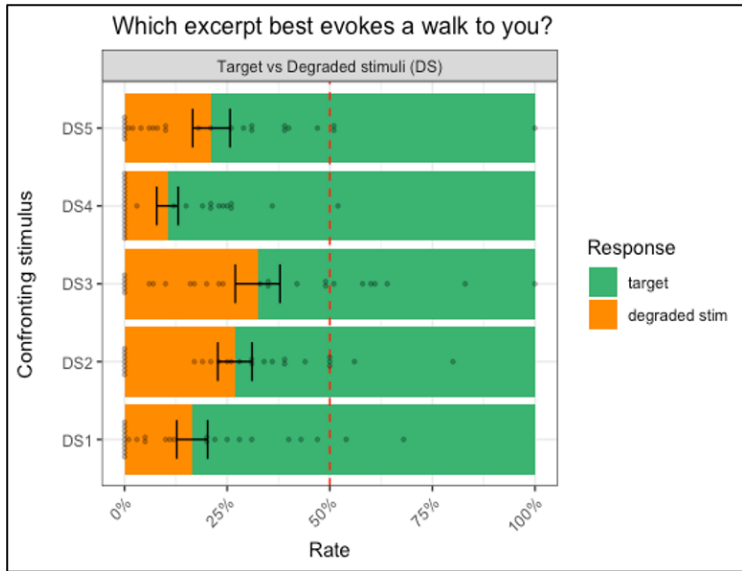
¹³A clarification is needed at this point: by ‘binarity’, we here refer to the metrical structure of the stimulus, i.e. that each bar here contains an even number of beats (4). Despite the presence of four separate musical events in our stimulus, the harmonic structure can be reduced to only two chords; hence breaking binarity does not entail breaking the pairs of the form $\langle \textit{bass}, \textit{chord} \rangle$ but pairs of the form $\langle \textit{chord}_1, \textit{chord}_2 \rangle$, i.e. $\langle \langle \textit{bass}_1, \textit{chord}_1 \rangle, \langle \textit{bass}_2, \textit{chord}_2 \rangle \rangle$.

Name	Condition violated	Rewriting procedure	Score	Change in the semantics*
DS 1	steadiness	Steadiness was broken by inserting rests so the time interval between musical events was not constant anymore and the metrical structure not identifiable.		Destroying regularity creates a sense of imbalance or hesitation.
DS 2	alternation	The back and forth from C to G was removed from the bass line and only C was kept, leading to the repetition of a same bass note and a chord 8 times in a row.		Under a 'walk' interpretation: the character seems stuck. Under a more abstract interpretation: an event is repeated and seems prevented from evolving.
DS 3	binarity	Steadiness and alternation don't constrain the number of musical events involved. Binarity is naturally what can represent a human walk the best as two feet and hence two steps only are involved. We thus replaced the binary alternation from C to G with a ternary alternation from C to G and then to E. E was selected because it is the only remaining available note from the C major perfect chord; hence the introduction of E as a third bass note would not create any harmonic novelty in the stimulus.		The introduction of a third component in the binary structure of the stimulus creates a ternary cycle incompatible with a walk.
DS 4	consonance	In the target stimulus, the bass line only contained singles notes and the chords were perfect chords: everything thus sounded extremely consonant. Here, bass notes were replaced by dissonant clusters of notes matching the pitch change in the target on beats 1 and 3, while the C chord was replaced by a dissonant augmented fourth interval (present in the cluster, as the chord on beats 2 and 4 in the target was derived from C).		The high level of dissonance creates an impression of great instability.
DS 5	chord proximity	We increased tonal distance between the two chords by replacing the second inversed C chord by a major F# chord. Just as for DS4, we almost preserved the structure of the target stimulus except that the chord on beats 4 is now different from that on beat 2, but we favored harmonic consistency between the bass and the corresponding chord over repetition (the only reason for which the chord was repeated in the target stimulus was that the two bass notes pertained to a same C major chord).		The great tonal distance between the two chords creates a sense of imbalance between the events, or, under a 'walk' interpretation, between the two steps (as if the character was limping).

* The purpose of this column is just to give an idea of how 'different' from the target the degraded stimuli sounded. These judgments come from the authors' own intuitions, along with some feedback collected during the piloting phases of the project. They do not aim at establishing any rule for characterizing the stimuli and only provide plausible descriptions of the semantic effects resulting from the destruction of the target's main musical properties.

Figure 5: Degraded stimuli. Stimuli can be accessed at the following links: [DS1](#), [DS2](#), [DS3](#), [DS4](#), [DS5](#)

4.3 Results



Reminder: Degraded stimuli	
DS5	Breaks tonal proximity
DS4	Breaks consonance
DS3	Breaks binarity
DS2	Breaks alternation
DS1	Breaks steadiness

Figure 6: Preference for target over degraded stimuli

We found that the target stimulus did significantly better at evoking a walk in participants than any degraded stimulus.¹⁴ These results suggest that each of the five identified conditions did contribute to the association of the target stimulus with a walking event. A pairwise comparison of each possible pair of degraded stimuli revealed a few significant differences in the preference for the target depending on which degraded stimulus it was paired with. In particular, we found marginally significant differences between DS1 and DS3; DS4 and DS2; DS4 and DS3 respectively: even though all degraded stimuli were judged as being less evocative than the target stimulus, some of them were particularly less evocative than others. These differences might indicate some kind of hierarchy between the tested conditions. For instance, breaking steadiness appears to alter the inference significantly more than breaking binarity; and breaking consonance seemed to alter the inference significantly more than breaking the alternation condition or the condition on binarity. It is however too soon to state anything more about the exact contribution of each condition to the generation of the associated representations. Further work is required to make sure that some musical properties represent more fundamental properties of the denoted event. It seems however reasonable to posit that the musical properties that are interpreted in terms of the most

¹⁴The statistical details are provided in Appendix B.

essential properties of the evoked scene are the ones whose violation blocks the evocation of a walk the most.

5 Enriching the model of music semantics

After having (i) confirmed that the target stimulus was a good musical candidate for evoking a walk (Section 3) and (ii) identified and tested some properties responsible for this association (Section 4), we then wanted to check whether these results fitted within a model of music semantics such as that of Schlenker (2017). This model, which relies on Rules of Preservation of Orderings (RPOs), is presented in the next section. We will argue that our results reveal a limitation of this model, and we suggest ways to enrich it.

5.1 Ordering preservation-based model

We will first explain what rules of preservation of orderings are, and we will illustrate these rules with a single musical property (for clarity): harmony. In Schlenker (2017), harmonic stability is taken to be interpreted in terms of physical stability of objects through a rule of preservation of ordering (RPO):¹⁵

Let $M = \langle M_1, \dots, M_n \rangle$ be a musical event, and O an object involved in event $E = \langle e_1, \dots, e_n \rangle$. M is true of E if when M_i is less harmonically stable than M_k , O is in a less stable position in e_i than in e_k .

This rule is one of several in a model of musical truth, which involves other requirements on time structure and loudness. Here, we simplify the model and reduce it to its condition on harmonic stability for simplicity. A little detour through picture semantics will now be helpful to clarify this condition. Let us take an example from Schlenker (2017) in Figure 7.

¹⁵In the original piece, this rule is used to define musical truth, hence the rule provides conditions as to how a musical event can be ‘true of’ an event. While we take inspiration from this framework, we won’t be using the notion of truth here, and we will approximate the meaning of ‘is true of’ by ‘can evoke’.

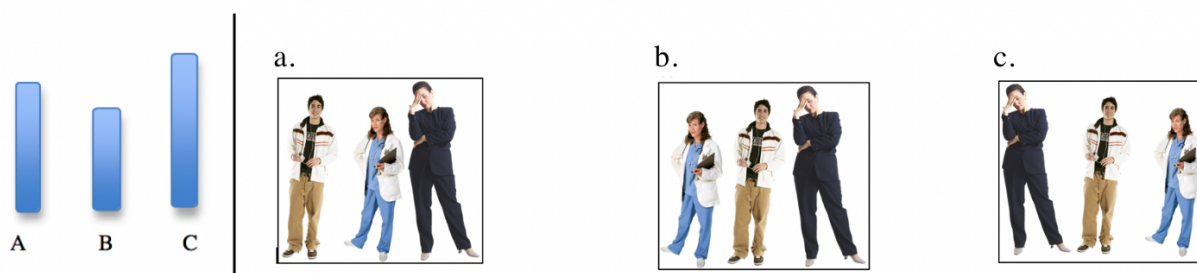


Figure 7: Preservation of ordering in pictures semantics

The question here is: which picture among pictures a, b and c is the graphical representation on the left true of? In other words, which picture is accurately represented by the graphical representation? Intuitively, picture a. is the best candidate: the graphical representation could denote many other situations, but the situation depicted in picture a. is one of them. Why is this so? Schlenker argues that this comes from the fact that the height of the bars in the graph is ordered in the very same way as the height of the characters in picture a. Roughly: $height(B) < height(A) < height(C)$. The semantic rule involved in making picture a. a possible denotation of the three bars on the left is thus a RPO with regards to the relevant property: in this case, height.

Musical properties are however of course not visually but auditorily perceived. What does this then mean that harmonic stability obeys the same kind of rule (namely, a RPO) as height does in the visual example above? It means that, if it is true that harmonic stability is interpreted as physical stability (just as the height of bars is interpreted as the height of people in the visual example), then the ordering in physical stability in a situation S must match the ordering in harmonic stability in music M for S to be a possible denotation of M . For instance, the succession of a first harmonically very stable chord and a second harmonically very unstable chord¹⁶ could denote any succession of two events whose physical stability evolves in the same direction, i.e. decreases.¹⁷ A series of two events such that the first one is someone standing and the second one is someone falling is a possible denotation, as long as the latter event is represented as being less stable than the former.

¹⁶We will come back to this notion of harmonic stability later.

¹⁷Note that in Schlenker's model, rules on harmony are enriched with rules on time and loudness, which we will put aside here for the sake of simplicity.

Because of reasons mentioned in Appendix A, we will not be using harmonic stability in our analysis and we will prefer two related musical properties: (i) consonance and (ii) chord proximity, for which we provide a semantic analysis largely inspired by Schlenker’s in Appendix A. We assume the rules derived from these analyses in the following sections.

5.2 Wrong predictions from RPOs

When applied to our target stimulus, a model based on RPOs appears to make wrong predictions. To show this, let us first consider the piece in Figure 8:

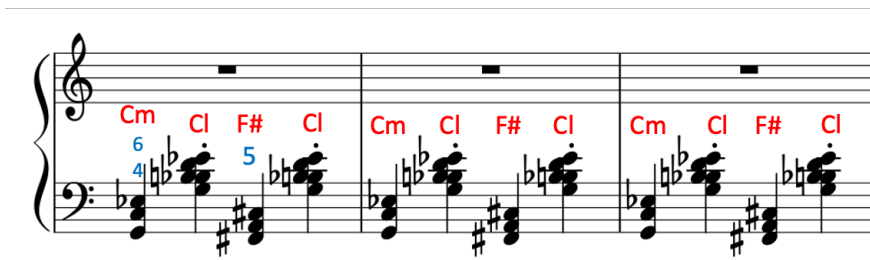


Figure 8: A piece which RPOs predict should evoke a walk

Given the evolution of consonance throughout this stimulus, and assuming RPOs on consonance as detailed in Appendix B, the stimulus should be able to evoke a walking event. Now, let’s consider our target stimulus once again:



Figure 9: Target stimulus

A similar formal analysis of the stimulus relying on RPOs, available in Appendix A, yields the same prediction: the target stimulus should be able to denote a walk.

As represented in Figure 10, both musical excerpts meet the criteria on consonance for denoting a walk.¹⁸ The prediction from RPOs is thus that both musical excerpts should be

¹⁸Formal details are available in Appendix A.

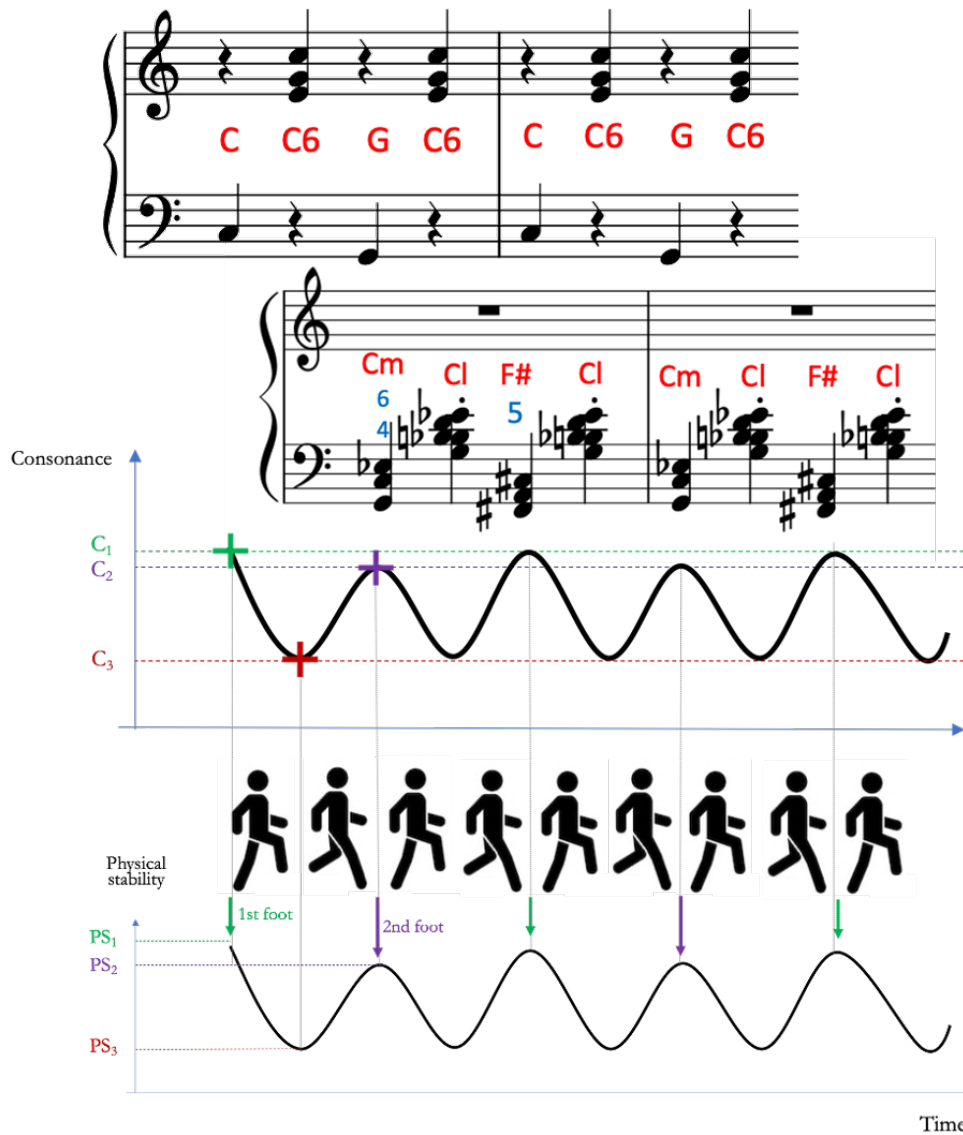


Figure 10: Consonance to physical stability mapping

evocative of a walk, and that no preference should be found for either stimulus. Yet, both excerpts are not as good at evoking a walk: the target stimulus is intuitively much better than the stimulus in Figure 8, which suggests that RPOs are insufficient to account for why the target stimulus evokes a walking event. We confirmed these intuitions experimentally and present the results in the following sections.

5.3 Stimuli

The stimuli consisted of the target stimulus, the excerpt presented in Figure 8 (referred to as ‘Ordering only 1’), and a second similar stimulus, shown in Figure 11 (referred to



Figure 11: Another piece which RPOs predict should evoke a walk

as ‘Ordering only 2’), which also seemed to fail to evoke a walk despite complying with RPOs on consonance. We introduced this second stimulus to make sure that our results were not excerpt-specific and could be generalized to any stimulus with the right consonance structure. This stimulus follows the exact same change in consonance as that in Figure 8, while suppressing the changes in range between successive chords (i.e. the alternation of lower-pitched and higher-pitched chords): the chords on beat 1 and 3 were simply shifted up by octave. The analysis we provide in Appendix A thus holds for this stimulus just as much: RPOs predict that the stimulus in 11 can evoke a walk. The target stimulus formed 2 pairs with each of these two stimuli complying with RPOs. Each participant was presented with the 2 pairs.

5.4 Results



Figure 12: Preference for target over stimuli satisfying ordering conditions

Since all three stimuli comply with RPOs, our prediction was that if music semantics only relied on RPOs, then all three excerpts should be associated with a walk with no distinction, and no preference for either stimulus would be found. However, the results show that the target stimulus was systematically and significantly preferred over any of the two other stimuli complying with RPOs.¹⁹ These results therefore suggest that we must reject the idea that RPOs alone account for the semantics of our target stimulus. We now provide two possible reasons for which we found that contrast, and suggest a way to enrich the model to capture the semantic differences we evidenced.

A first difference between the target stimulus and the stimuli in Figures 8 and 11 is that the former complies with the condition on relative chord proximity (all musical events in the target are either single tones from a major C chord, or a major C chord), while the two other stimuli do not: the Eb and F# chords on beats 1 and 3 are indeed very distant from one another in the tonal pitch space. This suggests that the similarity between the two steps²⁰ had to be represented by two musical events that are close to one another in the tonal pitch space. In other words, chord proximity seems to be a necessary condition because it is interpreted as the similarity between the two denoted events (the two steps), i.e. in terms of gait symmetry. This is consistent with our results from Section 4 in which decreasing chord proximity resulted in an impression of ‘limping.’

Second, another main difference between the target stimulus and the two other stimuli is that all musical events in the target stimulus are *absolutely* consonant, while in the two other stimuli, the introduction of the very dissonant clusters on beats 2 and 4 seem to block any evocation of a walking event. It seems that the slight differences in stability occurring during a walking event (for instance, the moment where the foot hits the grounds does not have the same stability as the moment where the foot is lifted) cannot be represented by *any* difference in consonance. First, a *small* difference in physical stability (such as the slight asymmetry between the two feet) should probably be represented by a *small* difference in consonance (such as that between a chord in its fundamental state, and an inverse chord 6/4). But this

¹⁹Details are provided in Appendix B

²⁰In Appendix B, we show that although both steps could be represented as having the same level of physical stability, footedness creates a slight asymmetry in stability based on which foot one is standing on.

is not enough: it seems that *all* events need to be denoted by a consonant musical event for the musical sequence to evoke a walk. Crucially, RPOs do not make this prediction: as long as a difference in consonance in a musical sequence M maps onto a difference in physical stability in an event E , RPOs predict that M can evoke E , regardless of the amplitude of this difference, and regardless of the correspondence from single stable events to single consonant musical events. Together with the results from Section 4, which established that suppressing overall consonance for all events blocked the association with a walk, this new piece of evidence gives extra-weight to an analysis under which absolute levels of consonance are interpreted as absolute physical stability of events.

6 Are all musical events semantically interpreted?

The previous sections established that (i) our target stimulus was more associated with a walking event than control stimuli, (ii) that the 5 conditions on this stimulus we identified reinforced a preference for the target stimulus when violated and (iii) that rules of preservation of orderings are insufficient to account for the association of our target stimulus and a character walking. In this last section, we present a final contribution of the present study through an investigation of the level of granularity at which this musical stimulus is semantically interpreted.

6.1 Motivation

Formal accounts of musical syntax have shown that musical structure can be represented at different hierarchical levels. One of the main contributions of the *Generative theory of tonal music* (GTTM) by Lerdahl and Jackendoff (1983) was to introduce the notion of *reduction*, which refers to the idea that listeners represent a complex musical piece at different levels of complexity, and accounts for musical intuitions about which musical events are more ‘important’ in a musical piece. GTTM introduces two such systems: time-span reductions, and prolongational reductions. Time-span reductions generate tree-like hierarchical structures mainly based on rhythmic stability, with hierarchical heads at each structural level that

correspond to the most structurally important musical event in that level. Prolongational reductions also generate tree-like structures but based on tension and relaxation patterns (i.e. on the relative stability of events), which patterns are themselves determined by stability conditions on branching, pitch, melody and harmony.

Schlenker (2017) provides a model of the semantic interpretation of headed events from time-span structures. This model introduces a notion of difference in importance between denoted events: roughly, headed (i.e. structurally important) events must denote events that are more important than the events denoted by non-headed events.²¹ This presupposes that both types of events are semantically interpreted, a non-trivial presupposition discussed in great detail in Zaradkzi (2021). Here, we tested an alternative hypothesis that not all musical events are interpreted as events, and, more specifically, that it is possible that less structurally important musical events (non-heads) are not interpreted as events but merely as ‘modifiers’ operating at a larger scale on events or groups of events. We tested this hypothesis experimentally using the stimulus presented in the next section.

6.2 Stimulus



Figure 13: The target stimulus enriched with an extra-8th chord on beats 2 and 4 of each bar, accessible [here](#).

The stimulus in Figure 13 was built from the target stimulus in Figure 1, in which the already present C major chord was repeated on beats 2 and 4, resulting in two 8th C chords on these two beats in each bar. From a structural perspective, it is reducible to the target stimulus through a time-span reduction: since each new chord is subordinate to the previous one, it is not represented at any higher structural level and in particular not at the immediately higher level which is that of the target stimulus, as represented in Figure 14.

²¹Note that Schlenker writes that this notion of importance should be clarified; and we agree.

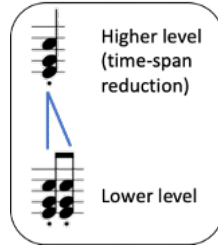


Figure 14: time-span reduction

Introducing this extra musical event in a recurrent way intuitively resulted in the stimulus sounding a bit happier or light-hearted, but without the clear impression that any event had been added to the denoted situation. In other words, it seemed to us that this new stimulus could denote the same kind of situations as the target stimulus did, suggesting that the addition of new musical events does not entail the addition of new denoted events.

We thus presented participants with a pair of stimuli made of the target stimulus and the enriched version with the additional chord in Figure 14 to check which of the two they would judge as denoting a walking event better. Our prediction was that if the stimulus was interpreted at a hierarchically higher level with less rhythmic complexity (i.e. the additional chord does not substantially modify the set of situations the piece can evoke), then no preference would be found for either stimulus. By contrast, if all musical events are interpreted as events (at the lower level of musical structure), then introducing a new chord should block the association with a walking event because this extra-chord would not correspond to any sub-event involved in the walking event; hence a significant preference for the target would be found. We also had participants compare this stimulus with previous control stimuli and degraded stimuli for more exploratory purposes; but a prediction was that if it were interpreted just as the target stimulus, then we would find a significant preference for it compared to any other type of stimulus.

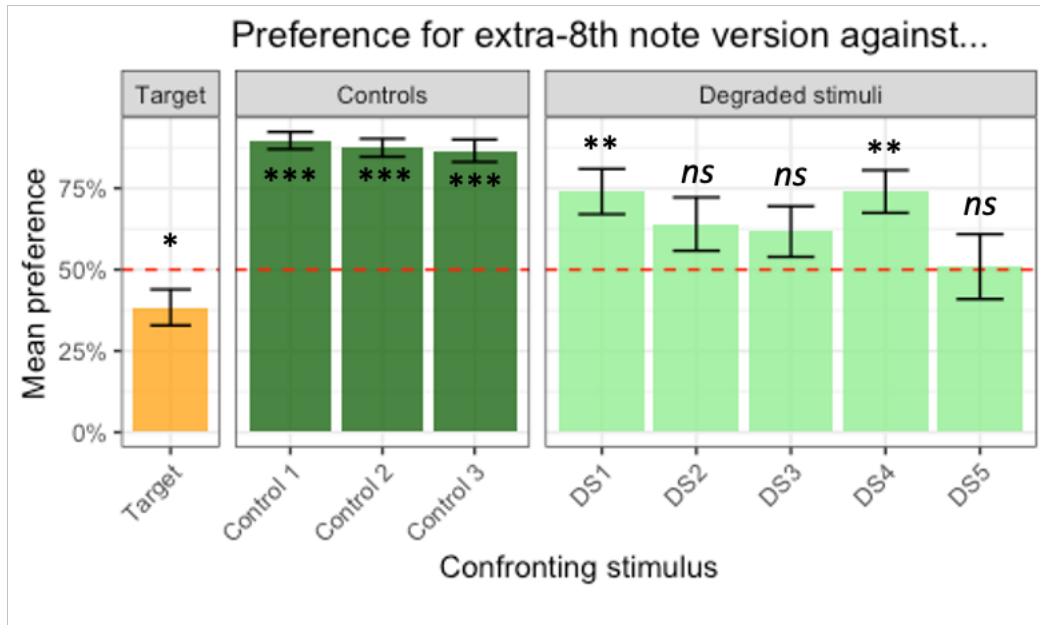


Figure 15: Preference for extra-8th chord version over the target stimulus, control stimuli and degraded stimuli

6.3 Results

The results in Figure 15 did not confirm our predictions: the enriched version of the target stimulus was judged as being evocative of a walk significantly less than the target stimulus.²² One must note that, qualitatively at least, the contrast found in endorsement remained small.

The enriched stimulus was very highly preferred compared to the three pieces used as control stimuli. This is not surprising, as that stimulus sounded much more similar to the target stimulus than to any of the control stimuli. The results from the comparison of the enriched stimulus and the degraded stimuli show that the former was only significantly more associated with a walking event than DS1 (breaking steadiness) and DS4 (breaking consonance). This suggests that introducing a new musical event without changing the fundamental structural properties of the target stimulus was perceived as altering the representation less than having two dissonant chords, and less than having an irregular sequence. This might suggest that these two conditions correspond to more fundamental properties of a walking event.

Together, these results indicate that introducing a higher hierarchical level (by means of doubling a musical event reducible to a more structurally important one) alters the association with a walking event *less* than violating some of the 5 conditions we identified. It was however

²²The statistical details are provided in Appendix B.

not the case that listeners can simply disregard a new musical event, even if that musical event was structurally less important.

7 General discussion and conclusions

We have presented an experimental paradigm investigating a case study in music semantics: the case of music evoking walks. We here summarize the main findings of this research and highlight some possible extensions.

First, our results give empirical support to the claim that music can be interpreted as evoking non-musical events. Second, it shows that some specific musical properties are responsible for evoking a walk, and we proposed that this was due to the form of our stimulus being reducible to a steady binary alternation of consonant and tonally close chords. Third, and drawing on a model developed in Schlenker (2017), we showed that rules of preservation of orderings, which state that what it takes for a music to evoke an event is that musical properties satisfy the same ordering as the interpreted properties, do not account for the evocation of a walk in listeners, and we proposed that listeners might be able to interpret some musical properties absolutely, and in particular interpret a consonant musical event as a stable physical event. Fourth, we tested the hypothesis that some structurally less important musical events might not be semantically interpreted, but our results suggest that this was not the case.

Another contribution of this work is to show that the semantic mappings from music to imagined scenes were independent of musical expertise. Each participant indeed had to self-report their level of musical expertise through a question displayed at the end of the experiment. We then checked whether there were differences in the responses based on musical expertise. A statistical analysis (detailed in Appendix B) revealed that this was not the case: musical expertise never had a significant effect on the preference for any stimulus. Although further work is required to test the effect of musical knowledge on a wider range of stimuli, these preliminary findings suggest that listeners are able to generate mental representations

from music regardless of whether they can consciously engage in an analysis of the structural properties of the piece.

Our methodology and experimental paradigm can finally be extended to the investigation of the semantics of any musical piece. In particular, this paradigm can be extended to investigate specific semantic effects from program music, such as music representing natural elements or animals, in which the semantic effects are strong. However, the difficulty in this task remains that any inquiry in music semantics requires a precise enough model of music semantics and a precise enough model of the *perception* or representation of the denoted event under investigation. While the former is developing, the latter might not be available for objects, situations or events traditionally represented by music. The case of walking, which has received a lot of attention in physiology and human motion analysis, was particularly well suited in our case. One could therefore privilege case-studies of musical sequences evoking events which have been modeled in the literature to make formal analyses of these events easier, and thus to make the formal analyses of the musical properties denoting these events easier as well.

References

- Cutting, J. E. and Proffitt, D. R. (1981), Gait Perception as an Example of How We May Perceive Events, *in* R. D. Walk and H. L. Pick, eds, 'Intersensory Perception and Sensory Integration', Springer US, Boston, MA, pp. 249–273.
- de Neeve, M. (2021), Larsons' musical forces in Schlenker's music semantics, Riga.
- Eitan, Z. (2013), How pitch and loudness shape musical space and motion, *in* 'The psychology of music in multimedia', Oxford University Press, New York, NY, US, pp. 165–191.
- Eitan, Z. and Granot, R. Y. (2006), 'How Music Moves', *Music Perception* **23**(3), 221–248.
- Eitan, Z. and Timmers, R. (2010), 'Beethoven's last piano sonata and those who follow crocodiles: Cross-domain mappings of auditory pitch in a musical context', *Cognition* **114**, 405–422. Place: Netherlands Publisher: Elsevier Science.
- Gabbard, C. and Itaya, M. (1996), 'Foot laterality in children, adolescents, and adults', *Laterality* **1**(3), 199–205.
- Gabrielsson, A. and Lindström, E. (2010), The role of structure in the musical expression of emotions, *in* 'Handbook of music and emotion: Theory, research, applications', Series in affective science, Oxford University Press, New York, NY, US, pp. 367–400.
- Hart, S. and Gabbard, C. (1997), 'Examining the stabilising characteristics of footedness', *Laterality* **2**(1), 17–26.
- Hausdorff, J. M. (2007), 'Gait dynamics, fractals and falls: Finding meaning in the stride-to-stride fluctuations of human walking', *Human Movement Science* **26**(4), 555–589.
- Hayafune, N., Hayafune, Y. and Jacob, H. A. C. (1999), 'Pressure and force distribution characteristics under the normal foot during the push-off phase in gait', *The Foot* **9**(2), 88–92.
- Helmoltz, H. v. (1877), 'On the sensations of tone as a physiological basis for the theory of music.', *Dover Publications*. .
- Huron, D. (2006), *Sweet Anticipation: Music and the Psychology of Expectation*, The MIT Press.
- Jackendoff, R. (2009), 'Parallels and Nonparallels between Language and Music', *Music Perception: An Interdisciplinary Journal* **26**(3), 195–204. Publisher: University of California Press.
- Koelsch, S. (2012), *Brain and music*, Wiley-Blackwell, Chichester, West Sussex ; Hoboken, NJ. OCLC: ocn767563922.
- Larson, S. (2012), *Musical forces: motion, metaphor, and meaning in music*, Musical meaning & interpretation, Indiana University Press, Bloomington. OCLC: ocn707212791.
- Lerdahl, F. (1988), 'Tonal Pitch Space', *Music Perception: An Interdisciplinary Journal* **5**(3), 315–349. Publisher: University of California Press.
- Lerdahl, F. and Jackendoff, R. (1983), *A generative theory of tonal music*, the MIT press, Cambridge, Mass. London.
- Meyer, L. B. (1956), *Emotion and meaning in music*, paperback ed., [nachdr.] edn, Univ. of Chicago Press, Chicago. Ill.
- Plomp, R. and Levelt, W. J. M. (1965), 'Tonal Consonance and Critical Bandwidth', *The Journal of the Acoustical Society of America* **38**(4), 548–560.

- Sadeghi, H., Allard, P., Prince, F. and Labelle, H. (2000), ‘Symmetry and limb dominance in able-bodied gait: a review’, *Gait & Posture* **12**(1), 34–45.
- Schlenker, P. (2017), ‘Outline of Music Semantics’, *Music Perception* **35**(1), 3–37.
- Sievers, B., Lee, C., Haslett, W. and Wheatley, T. (2019), ‘A multi-sensory code for emotional arousal’, *Proceedings of the Royal Society B: Biological Sciences* **286**(1906), 20190513.
- Todd, J. T. (1983), ‘Perception of gait.’, *Journal of Experimental Psychology: Human Perception and Performance* **9**(1), 31–42.
- Zaradzki, L. (2021), Les événements en sémantique linguistique et musicale, PhD dissertation, Université Paris Cité.

A Predictions from stimuli

The target stimulus used throughout the experiment was used because of its intuitive capacity to evoke a walk.²³ Still, there are theoretical reasons for which one could justify this stimulus based on (i) the mereological structure of a walking event, i.e. how subjects spontaneously represent or approximate a walking event hierarchically, (ii) the physical, physiological and mechanical analysis of walking and gait patterns from the literature, and (iii) a set of semantic rules that makes the correspondence between musical features and features of a walking event more explicit, i.e. how subjects convert musical information into a mental representation of a walking event.

Although other structures could be assigned to our perception of a walking event (i.e. other models could be considered for (i)), we show that the model we propose does not only make intuitive sense, but it is also consistent with the literature on gait physics. By combining this model with a semantic model from Schlenker (2017), we provide both experimental and theoretical reasons for which the stimulus evokes a walking event in the first place.

Modeling the hierarchical structure of a walking event

Although there exists an extensive literature on gait physiology and mechanics, there are only few studies on the perception of gait movement. Yet, we will rely on findings from Cutting and Proffitt (1981) and Todd (1983) to approximate the perceived structure of a walking event, along with general theories on how events are hierarchically and structurally perceived (e.g. Jackendoff (2009)). We are indeed interested in how music can evoke a walking event as perceived in listeners (not a walking event per se, regardless of how it is mentally represented in listeners). Let's thus first assume that a walking event can be represented as four successive subevents e_1 , e_2 , e_3 and e_4 as shown in Figure 16. We will then show how the stimulus used in the experiment matches this hypothetical mental representation of a walking event and how this mapping is grounded in the physics of walking.

Following intuitions by Jackendoff (2009), we assume that a walking event is, as any event, hierarchically structured: we represent events as successions of discrete sub-events, some of which are more salient than others and constitute 'heads', which less important sub-events depend on. Intuitively, any walking event involves the alternation of two steps: the first step can be roughly decomposed in a first event e_1 during which both feet are on the ground (which roughly corresponds to the stance phase in gait analysis), and a second event e_2 in which body weight is transferred onto one single foot (which roughly corresponds to the swing phase in gait analysis). The same pattern is then repeated with a third event e_3 with both feet on the ground, and e_4 with the second foot on the ground.

Modeling a walking event in terms of physical properties

But event structure is not all there is to a walking event, and we must now come up with a model of how physical properties evolve through a walking event, for we do not only want

²³Here, we borrow our methodology for selecting a specific stimulus from linguistics rather than from more standard approaches in cognitive psychology. In linguistics, it is indeed standard to rely on introspective judgments on the grammatically and/or the semantics of sentences to motivate research paradigms. Here, the stimulus plays the same role as any such sentence: the stimulus evoked a walk in the authors and informants prior to building the experimental design, and the remainder of the experiment tried to understand (i) how robust or systematic this effect was, and (ii) why it was or was not.

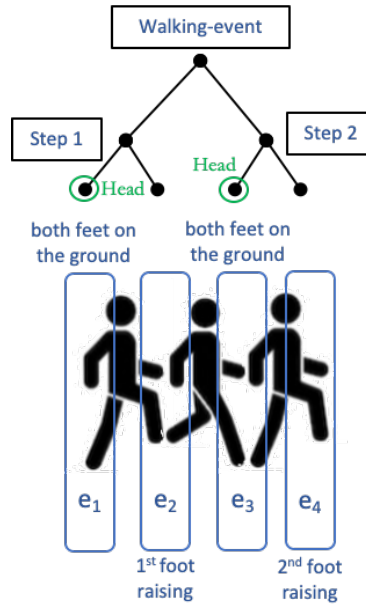


Figure 16: Tentative model of a walking event with tree-like structure

musical structure to map onto event structure, but we also want the interpretation of music to map onto our representation of the event. Following the literature on gait analysis, we use two descriptors of gait patterns: support and footedness.²⁴ Support refers to whether one foot or two feet bear the body weight, i.e. whether one foot or two feet are on the ground. Footedness refers to lower-limb laterality: just as handedness refers to the asymmetry in the use of hands, footedness is linked to foot preference in performing certain tasks such as climbing a step or kicking a ball. Both support and footedness result in differences in how body weight will be spread onto the floor surface, and eventually in how stable the event will be perceived. Intuitively, stability is higher when both feet are on the ground and lower when one foot is swinging (it is harder to stand on one foot than on two feet). Also, if the front foot is the dominant foot, stability will be slightly higher than if the non-dominant foot is leading. These intuitions are summarized in Table 17.

Event	e ₁	e ₂	e ₃	e ₄
Support	Double	Single	Double	Single
Front foot	Dominant	/	Non-dominant	/
Stability	++	-	+	-

Figure 17: Support, footedness and stability during a walking event

²⁴For several analysis of footedness, see Gabbard and Iteya (1996) and Hart and Gabbard (1997), and Sadeghi et al. (2000) for a review on limb dominance; See also Hausdorff (2007) and Hayafune et al. (1999) for detailed analyses of pressure and force distribution in gait.

Let us now formalize physical stability in a walking event. Let S_n be the physical stability of event e_n . We can now formally describe the evolution of stability of a walking event $E = \langle e_1, e_2, e_3, e_4 \rangle$ defined as the succession of four sub-events based on Figure 17:

(1) If E is a walking event, then $S_1 > S_3 > S_2 = S_4$.

We will now need to turn to the *musical properties* whose interpretation is likely to map onto physical stability. We will present two such properties in the following section: consonance, and chord proximity.

Deriving musical properties to evoke a walking event

The distinction between consonance and chord proximity is of particular importance with regards to how both properties are semantically interpreted. In the model proposed by Schlenker (2017), the corresponding musical property is harmonic stability, which comes with the following semantic rule: the more stable the musical event, the more stable the denoted event. However, the notion of harmonic stability has been disputed in both the musicological and psychological literature and seems to refer to a concept from music theory building on the notion of harmonic function, rather than a perceptual fact. Generally, harmonic stability refers to the harmonic function of a chord within the context of a given reference key (what Lerdahl (1988) refers to as a ‘tonal region’): in the key of C (the context, the tonal region), a G chord has the function of a dominant, which is commonly used as a way to both get away from the tonic and get back to it in standard harmonic progressions in Western tonal music. However, the concept of harmonic stability does have a psychological counterpart. Arguably, the harmonic stability of a chord in a given tonal context amounts to assessing the proximity that one chord entertains with the tonic chord establishing the context: a given chord C_1 is perceived as stable in the tonal context provided by a chord C_2 if C_1 is perceived to be closely related to C_2 . Crucially, chord proximity is not directly related to chord consonance, and most studies on chord proximity have only established the perceived psychological distance between consonant chords, for a good reason: the more dissonant the chords, the harder it gets to identify them in the tonal pitch space, hence the harder it gets to compute chord proximity. We will thus now tease apart these two notions, in particular because it seems that both properties can be semantically interpreted as physical stability, just as harmonic stability. We will first provide some information about consonance, before turning to chord proximity. We finally provide a semantic analyses of these two properties and derive the predictions for our stimuli.

Consonance

Consonance refers to the subjective experience of pleasantness when two (or more) simple or complex tones are heard at once. One of the most consensual observations about interval consonance is that consonant intervals tend to be such that the frequency ratio of the two tones can be expressed with small integers: the smaller the integers, the more consonant the interval. This relies on the core fact that tones have harmonics (also named overtones), i.e. musical sounds (complex tones) contain partials at frequencies that are different from the root’s. Helmholtz (1877) famously claimed that consonance depends on both (i) the difference in frequencies (aka interval width), and (ii) the pitch range of the constituent tones: a same interval is judged to be rougher in lower frequency ranges. As soon as two tones do not share a same fundamental frequency (i.e. the lowest partial), acoustic beats are created, i.e.

variations in perceived volume due to the periodic combination of the two sound waves with different frequencies, which results in an impression of roughness in the listener. The more complex the frequency ratio (i.e. the higher the integers), the more beats are created, hence the more dissonant the chord. Although dominant then and still now, this view was later updated, in particular by Plomp and Levelt (1965), who showed that the best predictor of whether an interval will be perceived as consonant or dissonant is actually not the interval width combined with the pitch range of each constituent tone, but the relationship with their critical bandwidth.

One must note that although this literature primarily focuses on the consonance and dissonance of intervals (made of two sounds), our experiment and the stimuli we tested involved both intervals and chords (made of more than two sounds; the word ‘triads’ being standard for three-tones chords). However, it was already acknowledged by Helmholtz that whatever theory accounts for interval consonance should generalize to chords consonance: the consonance of a chord is determined by the consonance of its constituent intervals. Typically, this generalization explains why only major and minor triads, along with their inversions, are considered consonant: those are the only ones that are made of the intervals which also sound consonant in isolation.

Chord proximity

Consonance can be said to be an inherent property of musical events such as intervals or chords: for instance, a listener can rate the consonance level of a single chord. However, it has been both taken for granted in music theory and shown in music psychology that listeners do have intuitions about how ‘alike’, ‘similar’ or ‘close to one another’ two chords sound, regardless of whether they are heard as consonant or dissonant. In other words, listeners are able to compute a certain ‘distance’ between two chords: this is the notion of chord proximity. Note that because this phenomenon only arises when two musical events are heard, it has to be ordinal in nature: one must always rate the proximity of at least two musical events. Among several theories available, Lerdahl’s *Tonal Pitch Space* is among the ones that has been able to capture most empirical findings in psychology, along with most assumptions assumed in classical Western tonal music theory (Lerdahl, 1988). Roughly, Lerdahl’s idea was that it is possible to predict the perceived distance or proximity between any two chords (either in a same tonal ‘region’, i.e. key, or from different regions) through a rule summing (i) the number of fifths separating the two chords on the standard circle of fifths (if the two chords are from a same region), (ii) the number of tones in common between the two chords and (iii) the number of fifths separating the two regions on the circle of fifths (if the two chords are from different tonal regions). Although very simple, this additive rule has proved surprisingly useful to account for proximity judgments of chords in listeners. In this paper, we therefore use Lerdahl’s theory from TPS as a baseline for characterizing chord proximity in our own stimuli.

Semantic interpretation of consonance and chord proximity

In Schlenker (2017), harmonic stability can be semantically interpreted in two ways, either in terms of (i) physical stability of the denoted object, event or situation and (ii) emotional stability of the listener. Examples from real music show that changing the harmonic structure of a piece results in this piece evoking situations with different stability patterns. For instance, going from a tonic chord (the most ‘stable’ one in a reference key) to a dominant chord (the

second most stable one in a same key) can denote an event that transitions from a stable state to a less stable state. Now that we have refined the underlying assumptions through the chord consonance/chord proximity distinction, we will generalize these assumptions on the interpretation of harmonic stability to chord consonance and chord proximity. We will assume that both consonance and proximity can be interpreted in terms of physical or emotional stability according to the following rules:

Property	Semantic rule
Harmonic function (Schlenker)	The more harmonically stable the musical event, the more stable the physical or emotional state of/in the denoted event
Consonance	The more consonant the musical event, the more stable the denoted event
Chord proximity	The closer two chords are in the tonal pitch space, the smaller the difference in stability level of the two denoted events

Figure 18: Possible rules for the interpretation of consonance and chord proximity

The reason for which we have introduced chord proximity might now become clearer: chord proximity captures cases where the notion of harmonic stability cannot be used. In particular, it states that it is possible to interpret chord proximity as a difference in stability in two denoted events when no tonal context is available, a prediction that a model based on harmonic function does not make. However, there are cases where despite the existence of a tonal context, the distance to the tonic remains constant through a stimulus: this could be said to be true of our target stimulus, which only involves musical events derived from a C chord. In this case, consonance helps predict which of two C chords will be perceived as more stable, based on their inversion state.

It is now possible to reconstruct the rule on harmonic stability from the rule on chord proximity by having one of the two musical events denote the tonic. It follows from the rule on chord proximity that

- (2) The closer a chord to the tonic in the tonal pitch space, the more similar its level of stability to that of the tonic.

Given that the tonic is associated with maximal stability, it follows from (2) that:

- (3) The further away a chord is from the tonic chord in the tonal pitch space (through a rule computing distance such as that of Lerdahl's in TPS), the less stable it is.

When a tonal context is available, the semantic rule can therefore be rewritten:

- (4) Let m_1 and m_2 be two musical events in a tonal context C , d_1 and d_2 the distance to the tonic of C in the tonal pitch space, e_1 and e_2 two events with levels of physical or emotional stability S_1 and S_2 , respectively. If $d_1 < d_2$, then $S_1 > S_2$

If no tonal context is available, then the rule has to be derived from consonance instead:

- (5) Let m_1 and m_2 be two musical events in a tonal context C , c_1 and c_2 their respective consonance, e_1 and e_2 two events with levels of physical or emotional stability S_1 and S_2 , respectively. If $c_1 < c_2$, then $S_1 < S_2$.²⁵

²⁵Note that the previous rule can also operate in combination with the first one when a tonal context is available.

Now, let us apply these rules to derive the predicted denotations of the stimulus discussed in Section 5:

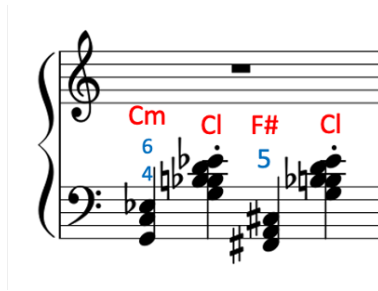


Figure 19: A stimulus which should refer to a walk based on RPOs

The music provides no tonal context. We will thus reason by limiting any interpretation in terms of stability to the interpretation of consonance.²⁶

Let our musical event be:

$M = \langle m_1, m_2, m_3, m_4 \rangle$
 with $m_1 = Cm(6/4)$, $m_2 = Cl$, $m_3 = F\#m(5)$ and $m_4 = Cl$

Let C_n be the consonance of event n .

Following our definition of consonance based on Helmholtz (1877), chords in their root position are perceived as more stable than inversed chords. So: $C_3 > C_1$.

A dissonant cluster is necessarily more dissonant than any perfect chord. It follows that: $C_3 > C_1 > C_2 = C_4$

A final twist is required at this point. Because the stimulus in Figure 19 was repeated, it is possible to have the stimulus start on m_3 ; hence m_3 is mapped onto e_1 , and m_1 is mapped onto m_3 . If we reassign indexes to take this into account, it follows that:

$$C_1 > C_3 > C_2 = C_4$$

We earlier defined a walking event as being such that:

(1) If E is a walking event, then $S_1 > S_3 > S_2 = S_4$.

We here see that the ordering in consonance of the stimulus in Figure 19 matches the ordering in stability of a walking event. So, RPOs predict that this stimulus can denote a walking event. Now, let us consider our target stimulus once again:

²⁶Note that this is an approximation used for simplicity and clarity in the model; it is for instance in this case possible to calculate the distance between chords on beats 1 and 3 based on Lerdahl's rules from TPS, and then simply consider that the cluster on beats 2 and 4 necessarily denotes a less stable event.

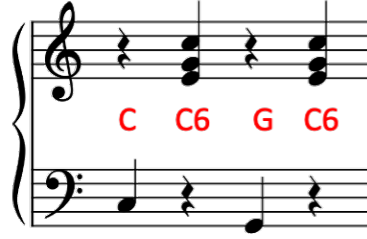


Figure 20: Target stimulus

Similarly, it is possible to compute the requirements on the physical stability of the denoted event of our target stimulus thanks to the very same rules. Let M be our musical event.

$$M = \langle m_1, m_2, m_3, m_4 \rangle \text{ with } m_1 = C, m_2 = C(6), m_3 = G, m_4 = C(6)$$

Let d_n be the distance of event n to the tonic C . Let S_n be the level of physical stability of a denoted event. Then $d_3 > d_1$ (G is further away from C than C from itself) and it follows from rule (4) that $S_3 < S_1$

Now, given that a single tone is necessarily more consonant than a 6 chord, it follows from rule (4) that:

$$C_2 = C_4 < C_3, \text{ therefore } S_2 = S_4 < S_3$$

$$C_2 = C_4 < C_1, \text{ therefore } S_2 = S_4 < S_1$$

$$\begin{cases} S_2 = S_4 < S_3 \\ S_2 = S_4 < S_1 \\ S_3 < S_1 \end{cases}$$

Therefore: $S_2 = S_4 < S_3 < S_1$

Hence, the ordering in consonance of the target stimulus matches the ordering in stability of a walking event. Rules of preservation of orderings on chord proximity and consonance thus predict that a walking event is a possible denotation for our target stimulus, just as it was for the stimulus in 19.²⁷

²⁷Another analysis of the target stimulus has been given by Zaradkzi (2021) reusing Schlenker's notion of harmonic stability in combination with Larson's theory of musical forces (Larson, 2012): The target stimulus is the repetition of a single bar made of four musical events: a single C , a C major chord in its first inversion state (6th), a bass G , and a C major chord in its 1st inversion state again. Harmonically, only one chord is present in the entire stimulus: the C major chord, made of the three notes C , E and G . It is usually acknowledged in music theory (and a sound perceptual hypothesis as well) that the most 'stable' form of a chord is its fundamental state. Here, the most stable event of the four has to be the bass C which is the tonic of the C scale. Comparatively, the bass G , which is the dominant of the same C scale, is usually considered to be a bit less stable (and often felt to be 'attracted' to the tonic for this very reason). Hence, in the key of C , in which the target stimulus is written, the bass C is a bit more stable than the bass G . Now, assessing the relative stability of the higher-pitch C chord requires an extra-step. We mentioned that this chord is in its 2st inversion state, which sounds less stable because of the presence of the 6th (compared to the presence of a 5th in the fundamental state, that corresponds to a frequency ratio of smaller integers hence perceived as more consonant). Although the perceived stability of that chord depends on the musical event that comes right before it, the fact that (i) it is a chord and not a single tone and (ii) is played in a higher register strongly supports that it should be heard as less stable (Schlenker, 2017; Larson, 2012). Consequently, the evolution of harmonic stability in

B Statistics

In this section, we present the detailed results of the statistical tests performed on the data to assess the significance of the contrasts in endorsement of the target stimulus and other stimulus types. For each pair, we performed Student's T-tests to assess whether the endorsement of the target stimulus was significantly above chance level, and report the corresponding t-scores and p-values. To check for statistical difference in endorsement in the confronting stimuli, we also performed pairwise comparisons of each pair of confronting stimuli, and report (i) the difference in endorsement, (ii) the lower and upper bounds of the 95% confidence interval, and (iii) the associated p-value.

Target *vs* control stimuli

Control	t-score	p-value
Control1	t (27*) = -14.984	< 0.001***
Control2	t (31) = -10.52	< 0.001***
Control3	t (29) = -9.0157	< 0.001***

*Note that the dfs differ because participants did not necessarily rate the same control stimuli.

Figure 21: Significance tests on the preference for the target stimulus over control stimuli

Pair	diff	lwr	upr	p adj
ctrl1-ctrl2	5.303571	-6.915831	17.52297	0.5569748
ctrl1-ctrl3	6.878571	-5.529455	19.28660	0.3869565
ctrl2-ctrl3	1.575000	-10.425203	13.57520	0.9474760

Figure 22: Pairwise comparisons of control stimuli

Target *vs* degraded stimuli

Degraded stimulus	t-score	p-value
DS1	t (26) = -8.9064	< 0.001***
DS2	t (25) = -5.5438	< 0.001***
DS3	t (25) = -3.2275	< 0.005**
DS4	t (28) = -15.146	< 0.001***
DS5	t (26) = -6.3371	< 0.001***

Figure 23: Significance tests on the preference for the target stimulus over degraded stimuli

the target stimulus follows the same pattern as in the excerpt shown in Fig. 19. For a further detailed formal analysis of how Larson's theory of musical forces can be integrated with Schlenker's theory, see de Neeve (2021).

Pair	diff	lwr	upr	p adj
PM2-PM1	10.403134	-6.0100208	26.8162886	0.4052957
PM3-PM1	15.941595	-0.4715593	32.3547502	0.0613849
PM4-PM1	-6.033206	-22.0080199	9.9416087	0.8339660
PM5-PM1	4.666667	-11.5909096	20.9242430	0.9319189
PM3-PM2	5.538462	-11.0288107	22.1057338	0.8868346
PM4-PM2	-16.436340	-32.5694593	-0.3032198	0.0435702*
PM5-PM2	-5.736467	-22.1496220	10.6766875	0.8695104
PM4-PM3	-21.974801	-38.1079208	-5.8416813	0.0022760**
PM5-PM3	-11.274929	-27.6880835	5.1382260	0.3223365
PM5-PM4	10.699872	-5.2749420	26.6746866	0.3480923

Figure 24: Pairwise comparisons of degraded stimuli

Target vs RPOs-only stimuli

Stimulus	t-score	p-value
Ordering only_1	t (44) = -6.9547	< 0.001***
Ordering only_2	t (44) = -8.9731	< 0.001***

Figure 25: Significance tests on the preference for the target stimulus over stimuli satisfying RPOs only

Pair	diff	lwr	upr	p adj
Ordering1/ordering2	-4.2	-14.97398	6.573976	0.4405926

Figure 26: Pairwise comparisons of stimuli complying with RPOs

Target vs Enriched version

Stimulus	t-score	p-value
Target	t (44) = -2.1008	0.04143*
Control 1	t (31) = 14.992	< 0.001***
Control 2	t (28) = 13.508	< 0.001***
Control 3	t (28) = 10.591	< 0.001***
DS1	t (12) = 3.4453	0.004847**
DS2	t (17) = 1.7053	0.1063 (ns)
DS3	t (19) = 1.5013	0.1497 (ns)
DS4	t (15) = 3.6575	0.002334**
DS5	t (19) = 0.090107	0.9291 (ns)

Figure 27: Significance tests on the preference for the target stimulus over the enriched version with additional 8th note.

Effect of musical training

In this section, we present the details of the statistical tests performed on the data testing for the effect of musical training. Musical training was assessed through direct self-assessment by participant and coded *via* a binary variable (0 for no training, 1 for some to a lot of training). For each category of stimuli, we ran two linear models on the data either including training as a predicting variable or not. The results we report in the table below were computed through an ANOVA performed between the two models. If training had an effect on the responses, then we expected the model containing training as a variable to fit the data significantly better than the other one. However, as seen in the table, no test was statistically significant, suggesting that musical training never significantly affected the responses participants gave for any category of stimulus.

Stimulus category	Sum of squares	p-value
Controls	-23.314	0.8957
Degraded stimuli	-72.012	0.8323
Ordering only	-2105.1	0.3012
Extra-8 th version	-9.3577	0.9382

Figure 28: Significance of the difference in predictions in two models (with or without musical training)

Chapter 5

Music-induced effects on body motion

Purpose

The previous chapter presented experimental data showing that listeners can represent a walking event from a musical sequence as long as that sequence meets certain conditions. In this chapter, we present an experimental paradigm exploring the extent to which musical properties affect real walking events (when listeners are walking). More specifically, we tested how pitch affects gait patterns, and found that pitch height correlates with properties of gait, which suggests that musical meaning might interact with the motor system.

CRACKING THE PITCH CODE OF MUSIC-MOTOR SYNCHRONIZATION USING DATA-DRIVEN METHODS *

Léo Migotti¹, Quentin Decultot², Pierre Grailhe², and Jean-Julien Aucouturier²

¹Institut Jean Nicod (CNRS - EHESS - ENS) - Département d'Etudes Cognitives, Ecole Normale Supérieure,
Paris, France - PSL Research University

²Department of Robotics and Automation - FEMTO-ST Institute (CNRS/Université de Bourgogne Franche
Comté), Besançon, France

ABSTRACT

The study of auditory-motor synchronization with music has been so far mostly concerned with timing. For instance, research has established that people are able to spontaneously coordinate with musical beat when walking or running (Styns et al., 2007). Yet, music is more than a metronome, and the relation between the spectral dimension of music, i.e. parameters such as its pitch, timber, or harmonic structure, and simultaneous motion remains nearly unknown. Here, we introduce a novel data-driven paradigm in which participants were asked to walk on a treadmill while listening to a large variety of musical tones systematically varied in pitch. Using analysis techniques inspired by psychophysical reverse correlation, we show that participants' gait patterns while walking to music spontaneously encode pitch height: despite being instructed to simply synchronize in time, participants steps were both longer and heavier on tones with lower pitches. These findings reveal that, similarly to time perception, pitch cognition is not purely 'disembodied' and suggest that listeners' spontaneous motor reactions to pitch might ground their ability to mentally represent music.

Keywords music | auditory-motor synchronization | pitch | walking | gait

* *Contributions:* LM, PG, JJA designed the experiment; QD and JJA developed the experimental apparatus and collected data; LM, QD, JJA analysed the data; LM and JJA wrote the manuscript, with contributions from PG.

Contents

1	Introduction	170
2	Results	172
3	Discussion	175
4	Materials and Methods	179
5	Acknowledgments	184
6	Bibliography	185

1 Introduction

Music in all human cultures is often experienced in motion, be it in dance, trance, military marches or lullabies (Blacking, 1974) and, even when we stand still, how we hear it remains strongly influenced by our motor system (Repp and Su, 2013). Because both music and movement unfold in time, the idea that there should exist a temporal mapping between the two appears intuitive, and indeed a vast body of research has shown that walking (Styns et al., 2007) and running speed (Van Dyck et al., 2015) spontaneously align with beat frequency (tempo). More generally, an extraordinarily rich literature has investigated people’s ability to move in synchronization with external auditory stimulation (Repp and Su, 2013; Thaut and Abiru, 2010), be it with classical finger tapping studies (Van Noorden and De Bruyn, 2009), limb movement (Torre and Delignières, 2008) or dancing (Toiviainen et al., 2010). Expressive timing in music performance, such as the final *ritardando* in the Baroque and Romantic period (Desain et al., 1996), also appear to share much of the temporal dynamics of motion (Sievers et al., 2013).

Yet, music is more than a sophisticated metronome and, while we know a lot about how music maps to motion ‘horizontally’ (in the time domain), little is known about whether the other ‘vertical’ (i.e. spectral) dimensions of music such as pitches, harmony and timbre relate to movement.

Indeed, while theories of embodied music cognition (Leman, 2007; Godøy and Leman, 2010) have been quick to extend beyond temporal/rhythmic parameters to propose that spectral aspects of music such as musical tension (Korsakova-Kreyn, 2018), pitch contours (Eitan and Granot, 2006) and consonance (Alves, 2012) also share cognitive representations with the physical features of body movement, empirical demonstrations of such links remain few and almost entirely limited to explicit judgements. For instance, in forced association tasks between music and dynamic visual stimuli, listeners tend to associate rising pitch contours with ascending motion (Eitan and Granot, 2006); random pitch intervals with jerky physical motion (Sievers et al., 2013); dissonant and unstable chords with motion having low physical stability (Schlenker, 2017, 2019); and alternations of consonant and harmonically-close chords

with the imagined situation of someone walking (the “Sorcerer’s Apprentice effect” (Migotti and Zaradzki, 2019)).

Such experiments demonstrate that music is able to evoke motion percepts or interpretations in listeners, which in itself is relatively uncontroversial (Tan and Kelly, 2004; Schlenker, 2017, 2019; Athanasopoulos and Moran, 2013) and is also reminiscent of well-studied pitch-space associations (Rusconi et al., 2006). It is however an entirely different question whether these associations are based on sensorimotor representations that are able to directly interact with the human motor system. For instance, if listeners in (Migotti and Zaradzki, 2019) reported less association to an imaginary walking situation for musical sequences including dissonant than consonant chords, does this mean that dissonant events would also disrupt their actual gait if they were to synchronize their steps to them?

In Komeilipoor et al. (2020), musical dissonance was found to alter the finger-tapping performance of participants asked to synchronize to either two dissonant or two consonant chords: in tasks where stimuli were dissonant, participants were less accurate in tapping, and remained less steady and regular after the sequence had stopped. However, in such a paradigm with a limited number of stimuli presented in separate blocks, it is difficult to conclude that it was the spectral/tonal properties of each individual sound event that matched and interfered with the corresponding motor event, or whether dissonance induced a more general cognitive or attentional load which interfered with the task on a broader and non-specifically-motor way (Bodner et al., 2007).

Here, we introduced a data-driven paradigm in which participants were asked to walk on a force-sensing treadmill while listening to a large variety of musical tones systematically varied in pitch. Such a paradigm has tremendous advantages for our research question. First, because each tone can be time-stamped to the X-Y-Z force time-series of the corresponding step, we were able to investigate how each individual sound event interferes with the corresponding motor event, similarly to the event-related methodology in neurophysiology. Second, because we can have participants step along with thousands of successive tones with randomly different pitch, the resulting data is amenable to analysis with psychophysical methods

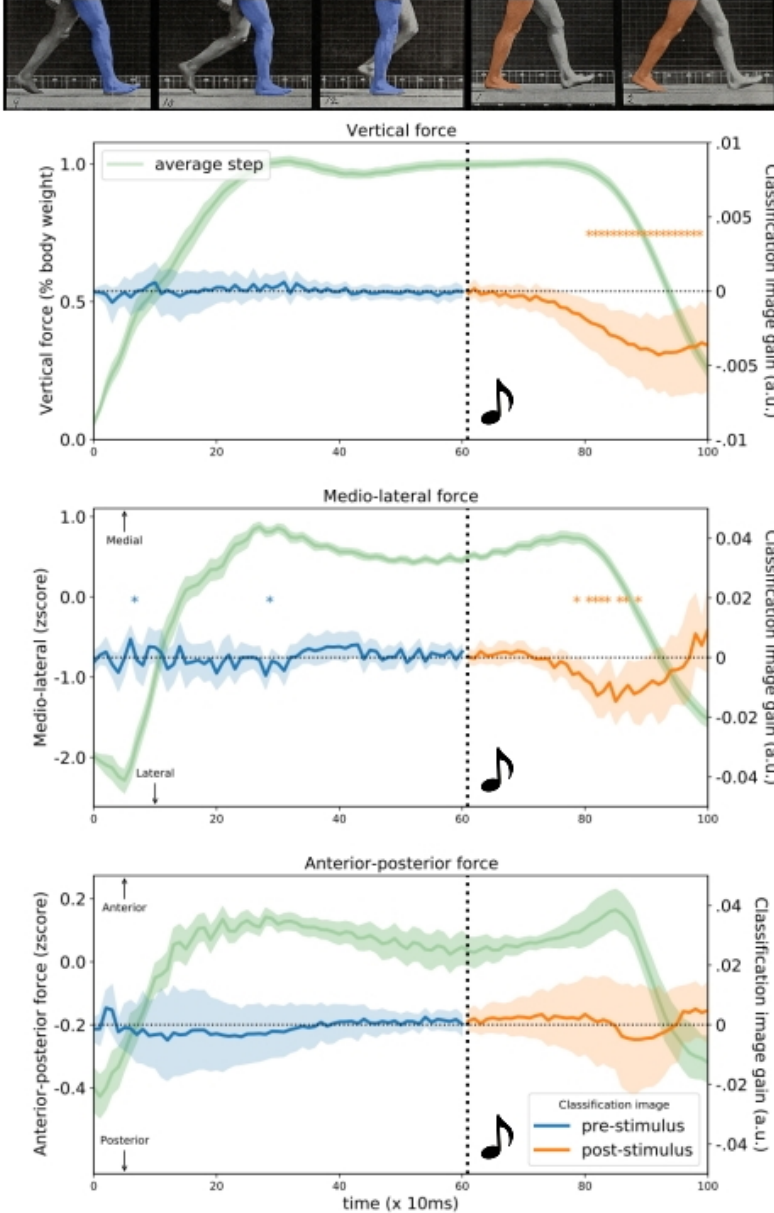


Figure 1: **Steps that occurred on lower-pitch notes exhibited more positive vertical force in their later, propulsive phase. Green:** Average vertical (top), medio-lateral (middle) and anteroposterior (bottom) force data for all participant steps ($M=668$ steps per participant). Participants’ step data displayed the expected saddle shape in the vertical and medio-lateral force dimensions. **Blue/orange:** Classification images indicating the direction of the statistical association between force at each time point and the z-scored, log-transformed pitch of the corresponding note, before (blue) and after (orange) note onset (marked by dashed line and note icon). Classification images showed a cluster of statistically-significant negative associations between pitch and vertical and medio-lateral force, post-onset, in the toe-off phase of the step.

inspired by reverse-correlation and we are able to infer what exact part of a participant’s steps is statistically influenced by pitch, in a purely data-driven manner.

2 Results

We let $N=20$ (female=9, male=11) young, Western, educated participants walk on a force-sensing treadmill while they listened to a 10min. sequence of musical tones (piano notes), whose pitch was randomly sampled within the two medium octaves of the western 12-tone chromatic scale (among 240 possible tones between $C4/255.8\text{Hz}$ and $B5/1045.8\text{Hz}$, sampled

at a frequency resolution of 10% of a semitone). Each successive tone in a sequence was repeated three consecutive times, at a rate of 75 events per minute (IOI=.8s). We then epoched the force time-series into individual steps and, in the manner of event-related analysis in electrophysiology, paired each step with its corresponding sound event. This resulted in an average M=668 trials per participant (see *Materials and Methods* for details). Six participants were removed because more than 20% of their steps were not synchronized with a musical event, leaving N=14 (female=6, male=8) for subsequent analysis.

Participants' step data displayed a saddle shape in the vertical and medio-lateral force dimensions and a progressive transition from anterior to posterior force, which is the expected shape for a stance including an initial, braking heel strike followed by a final propulsive 'toe off' (see e.g. Vaverka et al. (2015)). Step duration was stable at M=105ms (SD=71ms). Consistent with similar synchronization tasks in the literature (Repp and Su, 2013), we found that participants' steps steadily anticipated the onset of the notes, with a mean onset time occurring M=608ms (SD=83ms) after the start of support of the corresponding step, roughly at a local minimum of antero-posterior and latero-medial forces (Figure 1).

To analyze the relation between step force data and musical pitch, we used a data-driven technique inspired by the psychophysical method of "classification images" (Murray, 2011). Classification images indicate the strength of the statistical association between force at each time point and the z-scored, log-transformed pitch of the corresponding note: at each time point, we tested whether the images differed from zero with one-sample t-tests (see *Materials and Methods* for details).

Participants' classification images showed a cluster of statistically-significant negative associations between musical pitch and vertical force from t=800 to 1000ms ($t(13) \in [-2.20, -3.58]$, all $p_s < .046$; Figure 1), showing that steps that occurred on lower notes exhibited more positive vertical force in their later, propulsive phase. This cluster co-occurred with a similarly negative association in the latero-medial dimension (from t=780 to 880ms; $t(13) \in [-2.24, -3.71]$, all $p_s < .043$), indicating a medial direction consistent with the final toe off. There was no association between musical pitch and participants' steps in the antero-posterior dimension.

We then grouped each participant's trials and computed separated classification images for steps that occurred on the first, second and third repetition of each note. This revealed that the above effect was driven by negative associations that occurred specifically the second and third consecutive time any given note was heard and stepped onto. In the vertical force dimension, step 2 had a significant negative cluster from $t=710$ to 810ms ($t(13) \in [-2.23, -2.99]$, all $ps < .044$) and step 3 from 810 to 1060ms ($t(13) \in [-2.18, -4.01]$, all $ps < .049$), but none in step 1 (Figure 2-top).

This pattern of data is consistent with the interpretation that participants' steps performed on tones with lower pitch were heavier (showing more vertical force), but also that these steps were longer: because classification images were computed by aggregating steps of varied duration, it is possible that steps that had (positive) vertical force for longer amounts of time occurred more often for lower pitch, and that shorter steps, whose vertical force reached zero earlier, were conversely associated with higher pitch. For a confirmatory analysis, we therefore computed separate repeated-measure correlations (Bakdash and Marusich, 2017) between the (log-transformed) pitch in each participant's trials and the weight (computed as the mean vertical force between 800 and 1000ms) and duration of the corresponding steps: both correlations were significant at step 3 (weight: $r(2291)=-0.043 [-0.08, -0.0]$, $p=.03$; duration: $r(2291)=-0.066 [-0.11, -0.03]$, $p=.001$; Figure 3). In addition, to verify that musical pitch had an influence on step weight beyond the effect on step duration, we repeated the classification image analysis by normalizing all steps to 100% duration (figure 2-bottom). Classification images in the vertical dimension continued to exhibit a significant negative cluster at step 2 (from $t=670$ to 740ms , $t(13) \in [-2.23, -3.72]$, all $ps < .044$), and repeated-measure correlations between log pitch and weight in the same range ($700-750\text{ms}$) remained consistent, albeit non significant: $r(2286)=-0.034 [-0.08, 0.01]$, $p=.09$.

All in one, we therefore found converging evidence that participants' steps were *both* longer and heavier when they stepped on musical notes with lower pitch. The effect on vertical force when lowering pitch from B5 to C4 was $M=+3.6\% [3.05\%, 4.19\%]$, and the effect on duration was $M=+1.4\text{ms} [1.19\text{ms}, 1.61\text{ms}]$.

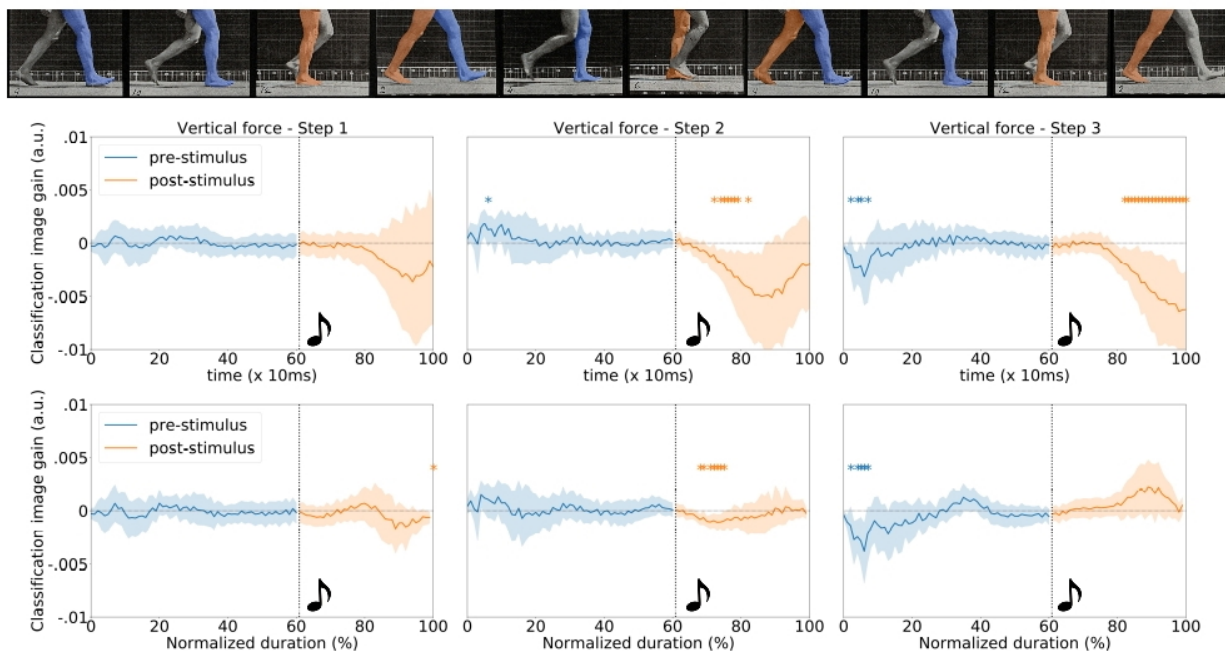


Figure 2: **Negative associations between vertical force and pitch that occurred specifically the second and third consecutive time any given note was heard and stepped onto, and remained even when the effect of step duration was normalized. Top:** Classification images between vertical force at each time point (actual duration, in ms) and the z-scored, log-transformed pitch of the corresponding note, before (blue) and after (orange) note onset; computed separately for the first, second and third consecutive repeat of each note. **Bottom:** Classification images in the same conditions, computed by normalizing all steps to 100% duration.

Finally, to test whether the influence of pitch was an effect of musical expertise, we tested for possible differences between the classification images of musically- and non-musically trained participants, using paired t-tests at each time point. There was no difference between both types of participants at any time point, for any of the effects we tested, regardless of whether we used participants' reported years of training or a standard test of musicianship (Gold-MSI) as a cutoff measure (e.g. vertical force, all p s $> .55$ from $t=80$ to 100 ms).

3 Discussion

To investigate whether musical pitch influences the gait parameters of simultaneous motion, we used a data-driven paradigm in which participants were asked to walk on a treadmill while listening to a large variety of musical tones systematically varied in pitch. Using analysis techniques inspired by psychophysical reverse correlation, we show that participants' gait patterns while walking to music spontaneously encode pitch height: despite being instructed

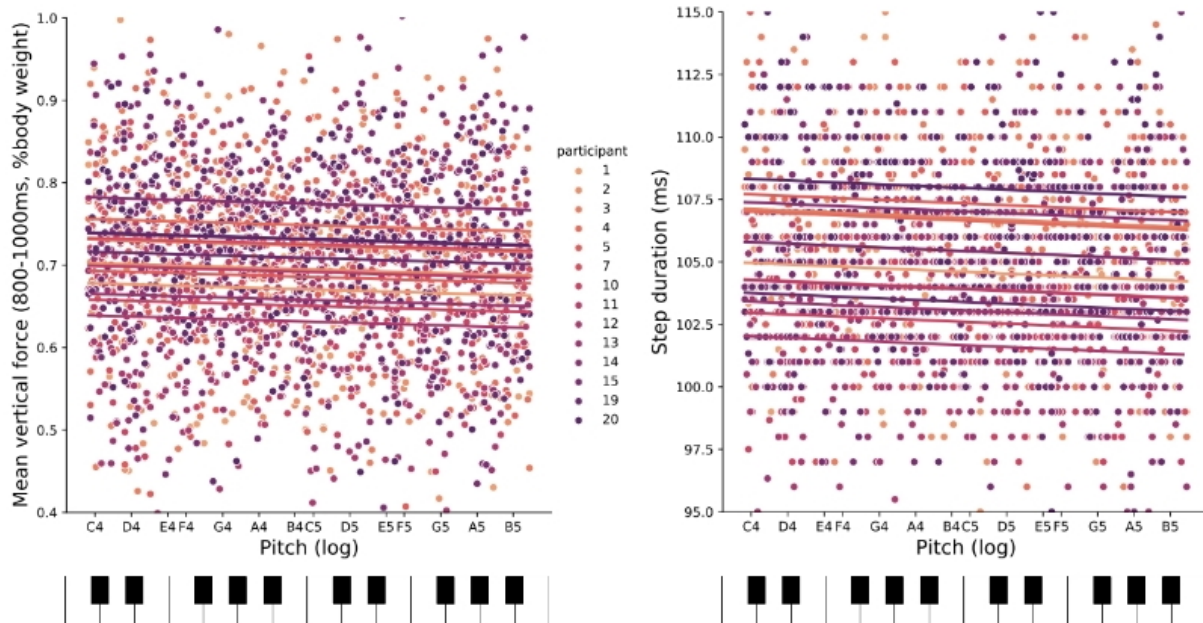


Figure 3: **Confirmatory analysis that participants’ steps performed on tones with lower pitch were heavier but also longer.** **Left:** Negative repeated-measure correlation between the (log-transformed) pitch in each participant’s trials and the toe-off weight (computed as the mean vertical force between 800 and 1000ms) of the corresponding step. **Right:** Negative repeated-measure correlation between the log-transformed pitch in each participant’s trials and the duration of the corresponding step.

to simply synchronize in time, participants steps were both longer and heavier on tones with lower pitches. This effect was demonstrated vertically and latero-medially in the toe-off stage of the steps (Figure 1); it occurred specifically the second and third consecutive times any given note was heard (Figure 2); and it was present similarly in musically- and non musically-trained participants.

Despite our analysis being entirely data-driven, the temporal extent of the effects found here (from 700 to 1000ms within step) was consistent with the constraints holding on the sensory-motor system. First, none of the effects occurred before the onset of the corresponding sound event, which was on average around 600ms within step. Second, the fact that the effect of a new note did not occur in the first time this pitch was heard is consistent with the sensory latency of processing musical pitch (Godey et al., 2001) and the motor latency of initiating walking motor commands (Sinkjær et al., 2000), which would collectively leave little time for the central nervous system to initiate any modulation of walking gait within the 100-400ms after the sound is first heard. In a comparable psychoacoustic study (Kadosh et al., 2008), motor reaction times (button press) measured in response to pitch deviations of

the range used here (2 to 13 semitones) were in the range of 750-1100ms post-onset. At a rate of 75bpm (IOI=800ms), such latencies would correspond to the range [-50ms, + 300ms] around the onset of the *next* consecutive note event, which is very similar to what we see here.

Our finding that our participants associate low pitch with longer and heavier motion is interesting to discuss in the context of the larger theoretical debate whether music can be said to have meaning (Leonard, 1956) and a semantics (Schlenker, 2017, 2019). Indeed, our results are reminiscent of a number of other studies reporting cognitive associations between musical pitch and other spatial concepts, such as elevation - the SMARC effect (Rusconi et al., 2006) - or speed (Broze and Huron, 2013); between falling pitch contours and descending motion (Eitan and Granot, 2006); between the repetition of low-pitched sound and quantity of motion in dance music (Van Dyck et al., 2013); between consonant and stable chords and motion stability (Schlenker, 2017, 2019); and between alternations of consonant and harmonically-close chords and the imagined situation of someone walking (Migotti and Zaradzki, 2019). That musical pitch can be interpreted as *representing* weight or stability therefore reinforces an emerging view that music is not only iconically referential, e.g. in the mapping between melodies and animal characters in Prokofiev's *Peter and the Wolf* (Trainor and Trehub, 1992), but can support a rich typology of semantic inferences about the properties and actions of non-auditory objects (Schlenker, 2019), in the same sense as language, gestures or visual animations (Tieu et al., 2019).

Even though the emergence of such representations is often studied in a musical context, one may question whether they are specifically musical, or of a more generic cognitive nature. Many existing cognitive mappings between, e.g., pitch and speed (Broze and Huron, 2013) are thought to result from the associative learning of events co-occurring in the natural environment (e.g. doppler effects (Mcbeath and Neuhoff, 2002)). For the behavior demonstrated here, it appears plausible that participants have developed sensory associations between heavy objects that hit the ground forcefully, and the resulting low or deep impact sounds, and that there is nothing intrinsically musical to this mapping. This suggests, as others have also proposed, that some of the semantics of music is “continuous” with generic

auditory cognition (Schlenker, 2017), in the same way that musical emotions are thought to be partly evoked by mimicking the acoustic properties of non-musical emotionally significant events, such as thunder (Ma and Thompson, 2015), or emotional vocalizations (Bedoya et al., 2021). In the present study, the fact that musically and non-musically trained participants showed similar effects shows that musical expertise is in any case not necessary for the association between pitch and stepping weight/duration to exist.

In comparison with previous work, we find remarkable that the present results did not emerge from self-report tasks asking participants to explicitly compare stimuli, but from physiological data in an implicit walking task where participants were not instructed to pay attention to the pitch quality of the musical events. This suggests that part of these associations do not only derive from ‘disembodied’ mental representations, but are able to recruit sensorimotor representations that directly interact with the human motor system, in our case modulate the force characteristics of participants’ gait. From the point of view of music semantics, the fact that musical pitch sequences should modulate overt walking behavior suggests that music might not only be able to refer to or denote external objects or events (Putman, 1987), but also internal events (i.e. happening in the listener’s body): if music has meaning, then maybe part of this meaning is about the listeners themselves. This is reminiscent of the question whether the smile-like spectral features of certain musical instruments can induce positive emotions and the imitation of a smile in listeners (Bedoya et al., 2021).

A related question concerns the cognitive ‘chronometry’ of the emergence of music-semantic representations relative to physiological responses: it is possible that pitch is first processed semantically as denoting heavier and longer steps, and that this representation then interacts with the (incidental) execution of the walking motor program; alternatively, pitch may first trigger an actual (or simulated) physical reaction, from which a mental representation is then lifted. Further experimental work, for instance with paradigms which examine stimulus evaluation while inhibiting or interfering with simultaneous motor behavior (Strack et al., 1988; Wagenmakers et al., 2016), will be needed to clarify the interaction of cognitive-evaluative and sensorimotor processes in the behavior exhibited here. Conversely, it is also possible

that motor actions such as stepping lightly or heavily can interfere with specific auditory perception tasks, in the same way that stepping on every second or third beat of a musical phrase can influence its perception as a march or a waltz (Phillips-Silver and Trainor, 2005). Finally, we believe our results may also motivate clinical applications where low-pitched sounds are used to stabilize gait in e.g. Parkinson’s patients (Hausdorff, 2007) or stroke survivors (Roerdink et al., 2007), possibly in conjunction with other well-described beneficial effects of tempo synchronization (Repp and Su, 2013).

Finally, if musical pitch has an effect on gait, it appears possible that more complex spectral properties (rich harmonic structure, timbre, consonance) do too. The present data-driven paradigm (for which we provide all experimental software as open-source²) therefore opens up a vast domain of research to investigate the interaction of each of these dimensions with human motion, and elucidate the intricate sensorimotor mechanisms that subserve our perception of such elusive multimodal percepts as motor and musical stability (Schlenker, 2021), predictive uncertainty (Hansen et al., 2021), tension (Costa and Nese, 2020) or smoothness/jerkiness (Sievers et al., 2013).

4 Materials and Methods

Participants

N=20 young (M=25.5, SD=7), Western, educated participants (male: 11, female: 9) participated in the experiment. All participants gave their informed consent prior to the experiment and were debriefed about the purpose of the research immediately after. Each of them received a 15 euros lunch voucher as a compensation.

N=13 participants (65%) self-declared as musicians (M=8.6y of musical training). In addition, all participants completed a short self-report survey consisting of 15 items extracted from the Goldsmith Music Sophistication Index (MSI) measuring their active engagement with music, perceptual abilities, musical training and singing abilities (Müllensiefen et al., 2014).

²<https://github.com/neuro-team-femto/treadmill>

Questionnaire scores significantly differed between musicians ($M=51.4$) and non-musicians ($M=38.8$; $t(11.7)=2.66$, $p=.0209$, $[2.26, 22.79]$, Cohen's $d=1.3$).

Procedure

We tasked our participants to walk on a force-sensing treadmill while listening to generative music constituted by a large variety of musical tones systematically varied in pitch. Prior to each session, participants were asked to wear comfortable shoes, and stood still once on each side of the treadmill for calibration. Treadmill was then started, its speed set at a slow walking speed ($M=2.5\text{km/h}$) and participants practiced walking to the sound of a metronome (75 bpm, period=0.8s) until they were comfortable with the task. Participants were then instructed to walk on the treadmill and to synchronize their steps with a sequence of piano notes they heard through headphones, without the support of handrails and as naturally as possible. Each session used an uninterrupted sequence of 750 sound events, played at a constant rate of 75bpm (period=0.8s), amounting to a total of 10min. Participants were given the cover story that we were interested in how well people could adjust their timing to the period of the sounds, and that the pitch content of the notes was irrelevant and just introduced for variety. They were instructed to focus their attention on walking synchronously with the sequence. After the session, participants were explained the true purpose of the experiment, namely that it was to examine the effect of pitch height on their walking gait.

Apparatus

Step data was acquired using a legacy instrumented treadmill (Tecmachine ADAL3D), equipped with two force platforms measuring the time series of forces of each limb during walking, along the X (medio-lateral), Y (antero-posterior) and Z (vertical) dimensions. The treadmill was interfaced to our software using a generic USB I/O acquisition card (NI-6008, National Instrument). Data was acquired at a sample rate of 1000Hz, and recorded continuously for the duration of one experiment (10min). While they walked on the treadmill, participants were presented with sequences of musical notes using commercial wireless headphones (Beats Solo3). Using custom software written in Python (<https://github.com/>

[neuro-team-femto/treadmill](#)), we synchronized step data acquisition with the triggering of each musical note, in order to be able to associate the time series of each step to the corresponding sound event.

Sound stimuli

Sound stimuli were extracted from the University of Iowa Musical Instrument Samples (MIS) dataset <https://theremin.music.uiowa.edu/MIS.html> and consisted in high-quality recordings of a Steinway & Sons model B piano, made in Nov. 2001 at the Voxman Auditorium, University of Iowa (Iowa City, USA) using a Neumann KM84 microphone. We selected one recording for each of the 24 notes spanning octaves 4 and 5, i.e. from C4 (263.3 Hertz, Hz) to B5 (1016.0 Hz), by steps of one semitone), played at medium dynamics (*mf*). To ensure that stimuli could be accurately synchronized to step data later in the experiment, we then normalized the loudness of the files, trimmed any initial silence (using a threshold at -30dB) and faded out each recording after 1.5sec so that all had the same duration.

We then produced a second derivative dataset by applying an algorithmic pitch transformation (pitch shifting) to each of the original files, in such a way that its fundamental frequency was reduced by 10, 20, 30, 40 or 50% of a semitone (or cents). Pitch shifting was done using a Python implementation of the phase vocoder algorithm (Burred et al., 2019). Finally, we ran an automated pitch analysis algorithm (SWIPE, (Camacho and Harris, 2008)) to document each musical note with its actual fundamental frequency, after pitch shifting. This resulted in a dataset of 240 musical notes, spanning the entire range of frequencies from a low-tuned C4 (255.8Hz) to a high-tuned B5 (1045.8Hz), by steps of 10% of a semitone (10 cents, corresponding to logarithmically increasing steps in Hz, from 1.5Hz at C4 to 5.9Hz at G5).

We then presented each participant with a random sequence of 750 sound events extracted from this dataset of 240 notes. Notes were played at the constant rate of 75bpm (period=0.8s), amounting to a total of 10min, and each note was repeated three consecutive times. In pilot studies, we experimented with several configurations, using only original recordings or using all pitch-shifted recordings, using notes from 2 octaves or 5 octaves, and using isolated or

repeated notes. Because we found that, consistently with the literature (Repp and Su, 2013), participants tended to anticipate the onset of the note to synchronize their steps (Figure 1), we decided to present each note three consecutive times so that steps 2 and 3 could be initiated and performed after one complete hearing of the corresponding note.

Step data signal processing

For each participant, we normalized the two time series of their left and right foot force data as percentage of the force corresponding to their standing weight, measured during calibration. We then resampled all time series at 100Hz and removed any linear trend across each session (due to treadmill electronics). Next, we segmented each participant's time series into individual steps using a simple threshold procedure: candidate start- and endpoints for each step were positioned where the vertical force time series crossed (resp. upwards or downwards) a threshold set at 5% of the amplitude range of the series, and we eliminated false positives that had a step duration shorter than 100ms. Finally, we associated each step with its corresponding sound event by selecting, for each step, the first sound whose time onset was between the step's start and end point. Steps for which no sound events were found (e.g. after the sequence ended) were deleted from the dataset, and steps for which more than one sound event were found were associated with the earliest of these events. The above procedure resulted in an average $M=668$ trials per participant (min=403, max=750, SD=114), each associating the time-series of a single step and the characteristics of the corresponding musical note. All analysis code (Python) is available on <https://github.com/neuro-team-femto/treadmill>.

Outlier selection

We selected as outliers participants who had more than 20% of their steps not associated with a musical event, on the rationale that this indicated a misunderstanding of the task, a low capacity to synchronize their walking gait to external sounds, or long attentional lapses. $N=6$ participants were removed from the dataset, leaving 14 participants for analysis (male: 8, female: 6).

Data-driven analysis

While previous work has focused on quantifying step data based on predefined characteristics (such as step and stride length, or heel and toe contact force (Vaverka et al., 2015)) and contrasting them along predefined stimulus properties (such as consonance (Komeilipoor et al., 2020)), we use here a data-driven strategy in which we let regions showing significant associations with sound emerge from an *a posteriori* analysis of participant responses to many, systematically-varied sounds.

In more details, we use a method inspired by the psychophysical technique of “classification images” (Murray, 2011). For each participant k , we compute a weighted time-series $\tilde{p}_k(t)$ by multiplying the time series of each the participant’s n steps $p_i(t)$ (with $t \in [0, 100]$ for a 1-sec step sampled at 100Hz) by a weighting factor z_i corresponding to the z-score of the log-transformed pitch of the corresponding note. We then average these weighted time-series over all n steps: $\tilde{p}_k(t) = \sum_{i=1}^n p_i^k(t) z_i$. Doing so, we negatively weight the patterns of step data that are associated with sound events with lower-than-average pitch, and positively weight those associated with high pitch. The resulting “classification image” $\tilde{p}_k(t)$ is itself a time-series analog to a single step, which represents the strength of statistical association between step force and musical pitch, as a function of time.

Note that for this analysis, it is important that the distribution of stimuli is non-skewed, so that classification images that do not differ from zero indicate no statistical association. Because musical notes are spaced logarithmically in Hz (lower notes are closer to one another than higher notes) the theoretical distribution of random stimuli is skewed towards low Hz, while it is uniform in log-Hz (or cents). For this reason, time-series data is weighted with the logarithm of the pitch (log-Hz), rather than in linear Hz, in the above calculation.

Ethics

The experiment was approved by the ethics evaluation committee of Inserm, the Institutional Review Board (IRB00003888, IORG0003254, FWA00005831) of the French Institute of medical research and Health, under the Opinion number 22-925.

5 Acknowledgments

This work was supported by ERC Grant StG 335536 CREAM (to JJA) and ERC Grant No 788077 Orisem (to P. Schlenker). Research was partly conducted at DEC, Ecole Normale Supérieure - PSL University. DEC is supported by grant FrontCog ANR-17-EURE-0017. Photographs of walking man featured in Figure 1 and 2 are adapted from public-domain photogravure by Eadweard Muybridge, 1887, Wellcome Collection gallery <https://wellcomecollection.org/works/v5t3b2ar>.

6 Bibliography

- Alves, B. (2012), ‘Consonance and dissonance in visual music’, *Organised Sound* **17**(2), 114–119.
- Athanasopoulos, G. and Moran, N. (2013), ‘Cross-cultural representations of musical shape’, **8**(3), 185.
- Bakdash, J. Z. and Marusich, L. R. (2017), ‘Repeated measures correlation’, *Frontiers in psychology* **8**, 456.
- Bedoya, D., Arias, P., Rachman, L., Liuni, M., Canonne, C., Goupil, L. and Aucouturier, J.-J. (2021), ‘Even violins can cry: specifically vocal emotional behaviours also drive the perception of emotions in non-vocal music’, *Philosophical Transactions of the Royal Society B* **376**(1840), 20200396.
- Blacking, J. (1974), *How musical is man?*, University of Washington Press.
- Bodner, E., Gilboa, A. and Amir, D. (2007), ‘The unexpected side-effects of dissonance’, *Psychology of Music* **35**(2), 286–305.
- Broze, Y. and Huron, D. (2013), ‘Is higher music faster? pitch–speed relationships in western compositions’, *Music Perception: An Interdisciplinary Journal* **31**(1), 19–31.
- Burred, J. J., Ponsot, E., Goupil, L., Liuni, M. and Aucouturier, J.-J. (2019), ‘Cleese: An open-source audio-transformation toolbox for data-driven experiments in speech and music cognition’, *PloS one* **14**(4), e0205943.
- Camacho, A. and Harris, J. G. (2008), ‘A sawtooth waveform inspired pitch estimator for speech and music’, *The Journal of the Acoustical Society of America* **124**(3), 1638–1652.
- Costa, M. and Nese, M. (2020), ‘Perceived tension, movement, and pleasantness in harmonic musical intervals and noises’, *Music Perception* **37**(4), 298–322.
- Desain, P., Honing, H. et al. (1996), Physical motion as a metaphor for timing in music: the final ritard, *in* ‘Proceedings of the International Computer Music Conference’, International Computer Music Association, pp. 458–460.
- Eitan, Z. and Granot, R. Y. (2006), ‘How music moves’, **23**(3), 221–248.
- Godey, B., Schwartz, D., De Graaf, J., Chauvel, P. and Liegeois-Chauvel, C. (2001), ‘Neuro-magnetic source localization of auditory evoked fields and intracerebral evoked potentials: a comparison of data in the same patients’, *Clinical neurophysiology* **112**(10), 1850–1859.
- Godøy, R. I. and Leman, M. (2010), ‘Musical gestures: sound, movement, and meaning’. OCLC: ocn298781501.
- Hansen, N. C., Kragness, H. E., Vuust, P., Trainor, L. and Pearce, M. T. (2021), ‘Predictive uncertainty underlies auditory boundary perception’, *Psychological science* **32**(9), 1416–1425.
- Hausdorff, J. M. (2007), ‘Gait dynamics, fractals and falls: Finding meaning in the stride-to-stride fluctuations of human walking’, **26**(4), 555–589.
- Kadosh, R. C., Brodsky, W., Levin, M. and Henik, A. (2008), ‘Mental representation: What can pitch tell us about the distance effect?’, *Cortex* **44**(4), 470–477.
- Komeilipoor, N., Rodger, M. W. M., Craig, C. M. and Cesari, P. (2020), ‘(dis-)harmony in movement: effects of musical dissonance on movement timing and form’, **233**(5), 1585–1595.
- Korsakova-Kreyn, M. (2018), ‘Two-level model of embodied cognition in music.’, *Psychomusicology: Music, Mind, and Brain* **28**(4), 240.

- Leman, M. (2007), *Embodied music cognition and mediation technology*, MIT press.
- Leonard, M. (1956), ‘Emotion and meaning in music’, *Chicago: University of Chicago* .
- Ma, W. and Thompson, W. F. (2015), ‘Human emotions track changes in the acoustic environment’, *Proceedings of the National Academy of Sciences* **112**(47), 14563–14568.
- Mcbeath, M. K. and Neuhoff, J. G. (2002), ‘The doppler effect is not what you think it is: Dramatic pitch change due to dynamic intensity change’, *Psychonomic bulletin & review* **9**(2), 306–313.
- Migotti, L. and Zaradzki, L. (2019), ‘Walk–denoting music: refining music semantics’, *Proceedings of the Amsterdam Colloquium 2019* .
- Müllensiefen, D., Gingras, B., Musil, J. and Stewart, L. (2014), ‘The musicality of non-musicians: An index for assessing musical sophistication in the general population’, *PloS one* **9**(2), e89642.
- Murray, R. F. (2011), ‘Classification images: A review’, *Journal of vision* **11**(5), 2–2.
- Phillips-Silver, J. and Trainor, L. J. (2005), ‘Feeling the beat: movement influences infant rhythm perception’, *Science* **308**(5727), 1430–1430.
- Putman, D. A. (1987), ‘Why instrumental music has no shame’, *The British Journal of Aesthetics* **27**(1), 55–61.
- Repp, B. H. and Su, Y.-H. (2013), ‘Sensorimotor synchronization: A review of recent research (2006–2012)’, *Psychonomic Bulletin & Review* **20**(3), 403–452.
- Roerdink, M., Lamoth, C. J., Kwakkel, G., van Wieringen, P. C. and Beek, P. J. (2007), ‘Gait coordination after stroke: Benefits of acoustically paced treadmill walking’, **87**(8), 1009–1022.
- Rusconi, E., Kwan, B., Giordano, B. L., Umiltà, C. and Butterworth, B. (2006), ‘Spatial representation of pitch height: the smarc effect’, *Cognition* **99**(2), 113–129.
- Schlenker, P. (2017), ‘Outline of music semantics’, *Music Perception* **35**(1), 3–37.
- Schlenker, P. (2019), ‘Prolegomena to music semantics’, *Review of Philosophy and Psychology* **10**(1), 35–111.
- Schlenker, P. (2021), ‘Musical meaning within super semantics’, *Linguistics and Philosophy* .
- Sievers, B., Polansky, L., Casey, M. and Wheatley, T. (2013), ‘Music and movement share a dynamic structure that supports universal expressions of emotion’, **110**(1), 70–75. Publisher: National Academy of Sciences Section: Social Sciences.
- Sinkjær, T., Andersen, J. B., Ladouceur, M., Christensen, L. O. and Nielsen, J. (2000), ‘Major role for sensory feedback in soleus emg activity in the stance phase of walking in man’, *The Journal of physiology* **523**(Pt 3), 817.
- Strack, F., Martin, L. L. and Stepper, S. (1988), ‘Inhibiting and facilitating conditions of the human smile: a nonobtrusive test of the facial feedback hypothesis.’, *Journal of personality and social psychology* **54**(5), 768.
- Styns, F., van Noorden, L., Moelants, D. and Leman, M. (2007), ‘Walking on music’, *Human Movement Science* **26**(5), 769–785.
- Tan, S.-L. and Kelly, M. E. (2004), ‘Graphic representations of short musical compositions’, **32**(2), 191–212.
- Thaut, M. H. and Abiru, M. (2010), ‘Rhythmic auditory stimulation in rehabilitation of movement disorders: A review of current research’, **27**(4), 263–269.

- Tieu, L., Schlenker, P. and Chemla, E. (2019), ‘Linguistic inferences without words’, *Proceedings of the National Academy of Sciences* **116**(20), 9796–9801.
- Toiviainen, P., Luck, G. and Thompson, M. R. (2010), ‘Embodied meter: Hierarchical eigenmodes in music-induced movement’, **28**(1), 59–70.
- Torre, K. and Delignières, D. (2008), ‘Distinct ways of timing movements in bimanual coordination tasks: Contribution of serial correlation analysis and implications for modeling’, **129**(2), 284–296.
- Trainor, L. J. and Trehub, S. E. (1992), ‘The development of referential meaning in music’, *Music Perception* **9**(4), 455–470.
- Van Dyck, E., Moelants, D., Demey, M., Deweppe, A., Coussement, P. and Leman, M. (2013), ‘The impact of the bass drum on human dance movement’, **30**(4), 349–359.
- Van Dyck, E., Moens, B., Buhmann, J., Demey, M., Coorevits, E., Dalla Bella, S. and Leman, M. (2015), ‘Spontaneous entrainment of running cadence to music tempo’, *Sports medicine-open* **1**(1), 1–14.
- Van Noorden, L. and De Bruyn, L. (2009), The development of synchronization skills of children 3 to 11 years old, *in* ‘Proceedings of ESCOM—7th Triennial Conference of the European Society for the Cognitive Sciences of Music. Jyväskylä, Finland: University of Jyväskylä’.
- Vaverka, F., Elfmark, M., Svoboda, Z., Janura, M. et al. (2015), ‘System of gait analysis based on ground reaction force assessment’, *Acta Gymnica* **45**(4), 187–193.
- Wagenmakers, E.-J., Beek, T., Dijkhoff, L., Gronau, Q. F., Acosta, A., Adams Jr, R., Albohn, D., Allard, E., Benning, S. D., Blouin-Hudon, E.-M. et al. (2016), ‘Registered replication report: strack, martin, & stepper (1988)’, *Perspectives on Psychological Science* **11**(6), 917–928.

Chapter 6

General discussion and conclusions

Contents

1	Main findings	190
1.1	Music conveys information about the extra-musical world	190
1.1.1	Music licenses inferences about an extra-musical reality	190
1.1.2	Music semantics obeys systematic rules	191
1.2	Musical representations interact with other cognitive domains	192
1.2.1	Musical meanings can be integrated to linguistic environments . .	193
1.2.2	Pitch height shapes gait patterns	193
1.2.3	Music semantics in music cognition	194
2	Limitations	195
2.1	Investigations restricted to multi-modal interactions	195
2.2	Simplified musical stimuli	196
2.3	An imperfect notion of meaning?	197
3	Avenues for future research	198
3.1	Exploring new semantic effects	198
3.2	The syntax/semantics interface	199
3.3	Considerations on aesthetics	201

We are reaching the end of our investigation of musical meaning. Throughout this dissertation, we presented several paradigms exploring a variety of music-induced effects. Because of the abstract nature of music semantics, and building upon the claim from Schlenker (2017) that “Semantic intuitions that would otherwise be very unclear can be sharpened by reducing the set of possible denotations,” we conducted research by having music interact with other systems such as language, visual perception, or the motor system. In this final chapter, we summarize the main contributions of the papers included in this dissertation in light of the theory of music semantics that we presented. We also provide a few insights regarding future avenues for research.

The diversity of the effects we reported supports at least two claims pertaining to different dimensions of music cognition. First, we showed that there is more to music and to musical representations than just reasoning about its form: music can also refer to an extra-musical reality through systematic rules, and we have argued that the set of inferences licensed about this musical reality can define musical meaning. Second, we showed that more general cognitive systems such as language or the motor system can process musical information and operate on musical mental representations.

1 Main findings

1.1 Music conveys information about the extra-musical world

Our results first show that music is able to convey information about an extra-musical reality in a rule-governed fashion. We first summarize the semantic associations which were experimentally confirmed, and then discuss the corresponding semantic rules involved.

1.1.1 Music licenses inferences about an extra-musical reality

Mostly consistent with previous findings and hypotheses on associations between music and movement (Eitan and Granot, 2006; Eitan and Timmers, 2010; Eitan, 2013; Schlenker, 2017), we confirmed that it is possible for listeners to associate different musical properties such as pitch, dynamics and timbre with designated properties of objects, be they visual objects in audiovisual animations, or imagined events such as a character

walking. In particular, we found that pitch and loudness can both be interpreted as the level of energy or the distance of a denoted object, while timbre is interpreted as the nature of an object, i.e. it is used for object identification in audiovisual scenes. From our study case on music evoking walks, we found that more fine-grained structural properties of music (such as steadiness and binarity) need to match the structure of the denoted event (in our case, a walk), and we also established that consonance can be interpreted as physical stability, while chord proximity (of two musical events) can be interpreted as a difference in stability between two denoted events. These findings therefore show that beyond musical properties whose semantics is lifted from normal auditory cognition (such as loudness and pitch), some music-specific properties (such as consonance and chord proximity) play a role in music semantics too.

1.1.2 Music semantics obeys systematic rules

We then showed that the preceding semantic associations obey systematic rules, and proposed three updates on these rules.

(i) While we confirmed that rules of preservation of orderings can account for the interpretation of pitch and loudness (i.e. musical properties and the corresponding object properties need to co-vary), we showed that they made wrong predictions as to which musical sequence could denote a walk. In particular, we provided evidence that consonance might be interpreted in terms of absolute levels of stability: a consonant musical event must denote a physical event, regardless of how its consonance level relates to that of other musical events.

(ii) We established that the following rule holds: When two objects are present in an audiovisual scene, the musical sound is paired with the object which satisfies the musical properties. This entails that it is not enough to state that *some* object satisfies the musical properties in the scene, but one and a same object has to satisfy these properties. This was particularly salient with loudness and timbre, which binds objects such that the object/timbre association must remain consistent throughout a scene.

(iii) We provided preliminary evidence that all musical events have to correspond to an

event in the denoted scene.¹

Together, our results show that it is possible for listeners to assess the compatibility of music with an-extra-musical reality by applying systematic rules of interpretation. We provide a summary of these results in Figure 1. As mentioned in the introduction, several theories had argued that musical meaning was only concerned with reasoning about musical form (i.e. musical meaning is only about processing musical structure, and draw inferences about this structure, for instance through anticipation (Huron, 2006)) or to musical emotions (Meyer, 1956). While our results do not entail that these theories are false, they suggest that musical meaning cannot be restricted to the kind of meaning these theories presuppose, and in particular not to musical form and representations of musical form. An exhaustive theory of musical meaning must account for the inferences music licenses about non-musical objects.

Property	Interpretation	Rule
Pitch	Distance	The higher the pitch, the closer the denoted object (marginally)
	Energy	The higher the pitch, the higher the energy of the denoted object
Dynamics (loudness)	Distance	The louder the music, the closer the denoted object
	Energy	The louder the music, the higher the energy of the denoted object
Timbre	Object identification	The timbre/object association needs to be consistent across the denoted scene
Consonance	Physical stability	<ol style="list-style-type: none"> 1. The more consonant the musical event, the more stable the denoted event/object (relative) 2. A consonant musical event must denote a physically stable event/object (absolute)
Chord proximity	Symmetry (in terms of physical stability)	The closer two musical events in the tonal pitch space, the smaller the difference in the physical stability of the denoted events/objects

Figure 1: Summary of the semantic rules associating musical properties with non-musical events

1.2 Musical representations interact with other cognitive domains

Besides establishing correlations with properties of objects, some of our results suggest that music triggers representations and behaviors that perhaps fall beyond the scope of

¹This rule relates to considerations on the syntax/semantics interface discussed in Section 3.2.

what can be captured by a music semantics (until further developments at least).

1.2.1 Musical meanings can be integrated to linguistic environments

First, and consistent with Tieu et al. (2019) on gestures and visual animations, we showed that when embedded in language, the information conveyed by musical stimuli can be divided into the different slots of the inferential typology. More specifically, we found that pro-speech musical gestures (musical stimuli replacing words in complex sentences) can give rise to (i) presuppositions (some of the informational content of music is taken for granted and unaffected by logical operations such as negation or question formation, while some of it is interpreted as being at-issue), (ii) scalar implicatures (it is possible to create musical stimuli with contrasted levels of logical informativity), (iii) supplements (musical stimuli can exhibit the same properties as non-restrictive relative clauses), and (iv) homogeneity inferences (musical stimuli can refer to a definite plurality of objects). We thus established that the mechanisms responsible for triggering different inferential types can process information which is neither linguistic in nature, nor visual.

1.2.2 Pitch height shapes gait patterns

Second, we showed that pitch height affects gait patterns in listeners engaged in walking, i.e. *spectral* properties produce physical effects on the body. While previous research had established time-related correlations between metric and rhythmic properties of music and walking patterns,² these can easily be accounted for by purely syntactic rules from musical structure to motion structure (time alignment for instance). The correlation we found between pitch and steps duration and force does not seem to be so easily accounted for by a purely syntactic account. In our case, pitch was physically interpreted in terms of properties of movement. Although not a semantic interpretation of pitch per se, this conversion of pitch into physical features obeys a systematic rule: a lower pitch is associated with a longer step with greater force. Interestingly, the physical reaction to changes in pitch (increase in stepping force, which can be associated to an increase in

²Among many examples: Styns et al. (2007); Moumdjian et al. (2019); Murata et al. (2017); Ready et al. (2019), see also Repp and Su (2013) for a review

weight) seems indirectly consistent with some hypotheses on the semantic interpretation of pitch as object size (lower-pitched sounds are associated with bigger objects (Eitan and Granot, 2006)), which suggests that semantic musical representations affect body motion.

Although very different in nature, these two findings from music embedded in language and music co-occurring with body motion supply additional evidence supporting the claim that musical information can be integrated with other systems. In other words: music can be taken as input in mental algorithms which are not music-specific, such as those generating the reported linguistic inferences, or those responsible for modulating the motor system.

1.2.3 Music semantics in music cognition

These findings however do not necessarily require an explanation in terms of music semantics (although they are consistent with it, under a few assumptions). It is for instance possible (although unlikely) that the meaning of musical stimuli is conventionally encoded, i.e. that listeners do not apply any particular rule to assign meaning to these stimuli which just happen to have been associated with certain things through exposure or learning. Similarly, the correlation we found between pitch height and gait properties might just be a low-level effect which does not imply any semantic representation.

It is therefore essential that any observation of a fact related to music cognition be followed by (i) a detailed analysis of the reasons why it is or is not relevant to try to account for that fact within a theory of musical meaning (all music-induced effects are potential candidates by default, but a theory needs to motivate the relation with music semantics), and (ii) a theory of how to integrate the relevant cognitive facts under investigation. For instance, Schlenker (2017) proposed an account of musical emotions in terms of a referential semantics, by positing that harmonic stability is interpreted as the emotional stability of an event (or object) or of the observer of that event. In the case of the correlations we found between pitch and gait, one should first ask: is it necessary to explain the role music semantics plays in this, and if so, how do we proceed? We can for

instance speculate that if music is able to refer to listeners themselves, then it becomes possible that they end up exhibiting motion properties inferred from the semantics of pitch. This is close to a point made in Schlenker (2017) about the possibility for music to be interpreted as *experienced* events. However, further theoretical work will first be needed to ground this hypothesis and derive the corresponding predictions.

2 Limitations

While we argued that most of our results fit nicely with a theory of music semantics, one could argue that the paradigms we developed and the stimuli we used raise questions as to (i) whether our findings generalize to music alone (typically when it does not interact with language, visual perception or body motion), and (ii) whether they generalize to real music.

2.1 Investigations restricted to multi-modal interactions

First, one might raise an objection related to the nature of the tasks involved in some of our experiments, namely that we never directly tested for spontaneous musical inferences in listeners. It is of course likely that participants would not have spontaneously interpreted our stimuli as someone walking, a white cube gaining energy, etc. But this is a much broader issue which applies to many cases of abstract representations, such as pictures or paintings. But even abstract visuals often seem more compatible with some scenes, and less with others. The problem with such abstract representations is that we do not necessarily have conscious access to their meaning, a problem that language typically does not have.

But precisely because music semantics is very abstract, one might argue that it is easier to ‘force’ meaning in music. This is a legitimate view: it is probable that someone who is forced to find meaning in something that does not have referential meaning tries to make sense of the object’s properties and starts making associations with other objects with some kind of systematicity. However, if listeners were just forcing interpretations, we would predict much weaker contrasts between stimuli complying with semantic rules

and stimuli not complying with semantic rules. That the contrasts we found can be explained by differences in productive rules generating meaning gives weight to the claim that these contrasts are not arbitrary. Although it is always possible that some non-semantic reasoning mechanism could approximate a semantic rule quite well, it remains unlikely that any music/world association results from such approximations.

2.2 Simplified musical stimuli

Second, one might argue that our findings are hardly replicable with real music because the stimuli we used were not musical enough. In other words, they might have lacked the structural complexity and sophistication of real music, hence a question of whether our findings generalize to cases of real music. But there seems to be no a priori reason to posit that the basic properties we worked on should behave differently: a crescendo should always be able to denote an increase in energy or a decrease in distance of an object to a viewpoint, unless these effects are canceled out by the interaction with other properties that we have not investigated in this dissertation. Any refinement of the theory, and in particular any additional hypothesis on the interpretation of a new property should therefore specify how that new feature is to be integrated with the properties for which a theory (and/or empirical data) already exists.

The use of simplified music might however have deeper implications. It is possible that the semantic mechanisms involved in the interpretation of minimal musical sounds are not quite the same as the ones involved when listeners are actively engaged in real music and listen to it *as such*. In other words, a potential risk of over-using these types of stimuli is to draw conclusions about the interpretation of auditory items that might not even be perceived as music. If that is the case, then it is not perfectly accurate to claim that we are testing for musical meaning, and it is more accurate to claim that we are testing for a more general class of meaning derived from non-musical sound properties. This is however not necessarily a problem, for at least some of music semantics is lifted from auditory cognition. Besides, some of our stimuli did have some reasonable musical structure (typically the ones used in our experiment on music evoking walks).

2.3 An imperfect notion of meaning?

The previous limitations might be seen as suggesting a limitation on the nature of the theory we have been relying on: that our definitions of ‘meaning’ and ‘semantics’ have been too permissive, or that we have used an ill-adapted notion of meaning throughout our investigation. This issue arises in other domains as well, where standard definitions of meaning cannot directly apply, such as in studies on animal communication, where meaning can be defined as a set of features of the contexts in which a given signal occurs (Berthet et al., 2023). In our case, one could argue that our notion of meaning is in particular different from a definition of meaning which entails intention, or a notion of ‘utterer’s meaning’ (Grice, 1957). Although there might be effects in music that are closer to reconstructing an intention (typically the composer’s or a narrator’s), none of our findings clearly showed this. We rather investigated things at a ‘pure’ semantic level, without considering pragmatic reasoning.

This begs the following question. Is it possible for a given representation (be it musical, pictorial, gestural, etc.) to denote something if there is no prior intention of conveying meaning? It is possible that, in the case of music, the answer depends on the type of music considered. For composed music, there is arguably an intention to convey information about the world to be reconstructed.³ For non-composed music, there seems to be different cases. In improvized music, there might still be an intention of conveying extra-musical information to be reconstructed (the performer’s). Although in other instances, such as computer-generated music, one could argue that no intention is present, a listener might still attribute an intention to a piece as long as it meets certain criteria. In any case, the results we reported here showed that both stimuli which were created with an intention of representing something (a character walking for example) and stimuli which were created with a less explicit intention (in the case of musical sounds simply satisfying structural constraints on loudness increase for instance) could give rise to semantic effects. It might then be that intention is irrelevant to musical meaning, at

³Whether the one that is reconstructed is actually the composer’s or a narrator’s, or to what extent the success of this reconstruction process conditions the aesthetics or the understanding of the piece, are interesting questions that we will leave aside for now.

least as we have chosen to explore it.

3 Avenues for future research

We have just seen that there is more to music than information about its form, and that musical properties can convey information about non-musical things. We will now argue that it is necessary to (i) evidence more semantic effects (and we suggest a few methodological strategies to do so), (ii) investigate how these relate to musical form through the exploration of the syntax/semantics interface, before providing (iii) some concluding remarks on aesthetics.

3.1 Exploring new semantic effects

Given the state of the field of music semantics, it is first necessary to document more semantic effects. From a methodological perspective, it seems that one can choose one of at least two strategies. First, through the exploration of the repertoire (and in particular of iconic programmatic music) and systematic controlled rewritings, it is possible to identify musical properties associated with semantic effects, and to isolate their semantics. Importantly, one should make sure to control for consistency when rewriting pieces: a rewritten musical excerpt should still comply with composition rules to avoid parasitic effects triggered by a violation of musical syntax. Just as is standard in linguistics, introspective judgments (to be confirmed by informants and possibly by experimental investigations, in particular in case the effects seem ambiguous) may prove very useful in exploring candidate semantic effects. Second, it might be that collecting experimental data from judgments on expressive music might help narrow down the investigation of new semantic effects to certain pieces and musical properties.

We will here provide two examples of musical properties to be explored in the future, which belong to two different categories of properties, each of which having some kind of homogeneity in the nature of the effects they trigger. We will not be discussing possible theories for which these properties convey the semantics we propose, and only provide these examples to show how the realm of properties to be investigated can be extended.

A first class of musical parameters that seem to convey strong semantic information are musical articulations. Those refer to how a musical event is being played by the instrument player through different techniques modifying its acoustic envelope. A clear example of how playing techniques affect musical meaning is the *pizzicato* technique, i.e. when bow-stringed instruments are plucked with fingers instead of being played *arco* (with a bow). Wherever it occurs, it seems like the pizzicato conveys some sense of lightheartedness, joy or playfulness.⁴ A second effect, of a more pragmatic nature, seems to arise especially when the pizzicato is used a lot: it then feels like the composer is being ironical. A second class of highly evocative parameters are those from which listeners seem to infer properties of the acoustic space. Reverb belongs to that class of parameters. In recent music and songs from Ethereal Wave and Dream Pop genres,⁵ the high level of reverb seem to be highly evocative of space and some kind of ‘boundlessness.’ A possible (speculative) semantic rule at this point could then be: the higher the level of reverb, the larger the inferred space. This would make sense under a theory in which listeners reuse some cognitive mechanisms from normal auditory cognition into music cognition.

As for any intuition about any semantic effect, both these suggestions should now be theoretically justified and then tested through introspective judgments first to make sure that the mappings we propose have some cognitive reality.

3.2 The syntax/semantics interface

Although we have already provided evidence in favor of some semantic effects found in music, referencing new ones will give weight to the claim that music can license inferences about extra-musical objects. But if music has both (i) a rule-governed syntax, i.e. rules

⁴While usually found only sparsely, some composers have had the audacity to go full pizzicato, such as Austrian composers Johann and Josef Strauss in their *Pizzicato Polka* from 1869., which Schubert (2004) describes the piece as ‘a short, lighthearted, almost comical piece’, which is very much in line with our own introspective judgments.

⁵These genres first developed in the late 1980’s with artists and bands such as American singer Julee Cruise (*Mysteries of Love* (1989)) and Scottish rockband Cocteau Twins (*I wear your ring* (1990), *Heaven or Las Vegas* (1990)), and later developed by incorporating more electronic features with, among many others, American indie rock band Yo La Tengo (*Our Way to Fall* (2000)), American solo group Grouper featuring Liz Harris (*Heavy Water I’d rather be sleeping* (2008)), and more recently American musical duo Beach House (*Myth* (2012), *Space Song* (2015)) or Icelandic alternative rock band Sigur Ros (*Ísjaki* (2013).)

on its form and (ii) a rule-governed semantics, i.e. rules on its meaning, then one must dig into how (i) relates to (ii).

For Schlenker (2017, 2019), the interface between syntax and semantics prompts two questions. First: Whenever two musical pieces are contrasted in terms of what they evoke in listeners, are those contrasts due to syntactic (structural) reasons, or to semantic reasons (related to the denoted objects)? Second: How does the syntactic structure affect the semantics (and if it does, is it the structure itself, or some more abstract mental construct based on this structure), and how does the semantics affect syntax in return?

Let us revisit some of our findings in light of these questions. Some of our results were indeed directly intended to investigate whether musical form was interpreted as it is, or whether a more abstract structure was rather interpreted. We tested some preliminary hypotheses on this matter in the case of music evoking walks. We found indeed that adding a hierarchically less important musical event to an excerpt evoking a walk significantly weakened the evocation of a walk. If our stimulus was interpreted at a higher level of musical structure, where this musical event merged into a hierarchically more important one and does not appear, then we should have found no contrast between the version containing this additional event, and the version which did not (because the two versions were then identical at this level). These preliminary results therefore suggested that our musical stimulus was interpreted at the lowest level of representation (where all musical events were represented). Further work is however needed to investigate other cases in more details, in particular by exploring more levels of representation of musical structure.

But if musical structure certainly determines musical meaning, it is also possible that the reverse is true, and that musical structure inherits some of its properties from music semantics, in particular ‘from the perceiver’s attempt to recover the structure of the denoted events’ (Schlenker, 2019). In other words: it can be that it is because music evokes an extra-musical reality with its own structural properties⁶ that listeners assign a given mental structure which aims at preserving (some of) the structure of the denoted

⁶Events have indeed been shown to have a hierarchical structure, see for instance Cotnoir and Varzi (2021) or Jackendoff (2009).

events. In our example: it is possible that the additional musical event was perceived as structurally less important because it denoted a less important event in the walk (for instance, a non-salient moment of the transition from one foot to the next).

These questions of the level of granularity at which a musical form is interpreted are at the core of the syntax/semantics interface. But these issues are however not specific to music. They arise just as much in pictorial representations. In this final section, we present an analogy between pictures and music, and explain how it may provide insights in terms of more general aesthetic considerations.

3.3 Considerations on aesthetics

The issues of granularity we mentioned call for a system of rules stating which level of the musical structure is interpreted, which should enrich the semantics that we have been developing, and it is possible that these rules are close to *marking rules*, which have semantic and aesthetic implications. Willats (1997) argues that artistic representational systems are essentially composed of pairs of (i) marks in the pictures, and (ii) features of the denotation that are marked.⁷ For instance, an artist might choose to mark only edges of a given denoted scene (say, a landscape), and to use black lines to represent these edges; that would be close to a very simplistic way of drawing. But the artist might well choose to do something entirely different, and mark only surfaces of a certain color, and use dotted surfaces to represent these surfaces. The very same theoretical framework might well apply to music: a full theory of music semantic must make these marking rules explicit (if any), and in particular clarify at which level(s) of musical structure are being marked.

Just as marking rules in pictorial representations seem to covary with pictorial styles (Willats, 1997), it might be that the analysis of marking rules in music proves useful in establishing a semantic-based typology of musical styles and systems of musical composition. Crucially, these marking rules may overlap with issues of granularity: some musical

⁷Importantly, these accounts of artistic representational systems are different from ‘pure’ pictorial representations, which Greenberg (2013, 2021) argued can be accounted for by a semantics in terms of geometrical projection onto a surface relative to a viewpoint.

system might require that all events be interpreted as world events, while some others might not. The question similarly applies to the level of granularity of the denotation: which properties of the denoted event have to be represented/marked?⁸

A final fundamental question bridges music semantics and music aesthetics in an even more direct fashion: how does musical content or meaning relate to aesthetic properties of music? In particular, a theory of music semantics needs to explain why listeners draw inferences from any kind of music, regardless of its aesthetic properties, and regardless of whether it is considered good or bad.⁹ It is possible that music semantics has implications in terms of aesthetic value, and that musical meaning (the set of denoted events) provides constraints as to both (i) what the best structure for a musical piece intended to represent a given event is, and (ii) what the best interpretation of a piece is (in the sense of a performance). This seems particularly relevant as there seems to exist a great consistency in what musicians rate as a good or bad interpretations (Geringer and Madsen, 1998); but further work on the interaction between music semantics and music aesthetics will be needed to explore the extent to which musical meaning contributes to aesthetic properties and aesthetic judgment of musical pieces and their interpretations.

In a nutshell, the expansion of music semantics will not only continue to give insights as to how music has been used, intentionally or not, to refer to and represent the world, but it will also prove very useful in addressing fundamental questions about how form relates to meaning and how meaning relates to aesthetic properties and judgments, in music and beyond.

⁸See Zaradkzi (2021) for a detail analysis of the notion of event in music semantics.

⁹As shown in Hesmondhalgh (2007), listeners have strong intuitions as to what qualifies as good or bad music. Washburne and Derno (2004) showed that these aesthetic judgments are often independent from aesthetic pleasure, in the sense that many listeners enjoy listening to music which they characterize as bad.

References

- Berthet, M., Coye, C., Dezechache, G., and Kuhn, J. (2023). Animal linguistics: a primer. *Biological Reviews*, 98(1):81–98.
- Cotnoir, A. J. and Varzi, A. C. (2021). *Mereology*. Oxford University Press, New York.
- Eitan, Z. (2013). How pitch and loudness shape musical space and motion. In *The psychology of music in multimedia*, pages 165–191. Oxford University Press, New York, NY, US.
- Eitan, Z. and Granot, R. Y. (2006). How Music Moves. *Music Perception*, 23(3):221–248.
- Eitan, Z. and Timmers, R. (2010). Beethoven’s last piano sonata and those who follow crocodiles: Cross-domain mappings of auditory pitch in a musical context. *Cognition*, 114:405–422. Place: Netherlands Publisher: Elsevier Science.
- Geringer, J. M. and Madsen, C. K. (1998). Musicians’ Ratings of Good versus Bad Vocal and String Performances. *Journal of Research in Music Education*, 46(4):522–534.
- Greenberg, G. (2013). Beyond Resemblance. *The Philosophical Review*, 122(2):215–287.
- Greenberg, G. (2021). Semantics of Pictorial Space. *Review of Philosophy and Psychology*, 12(4):847–887.
- Grice, H. P. (1957). Meaning. *The Philosophical Review*, 66(3):377.
- Hesmondhalgh, D. (2007). Audiences and Everyday Aesthetics: Talking About Good and Bad Music. *European Journal of Cultural Studies*, 10:507–527.
- Huron, D. (2006). *Sweet Anticipation: Music and the Psychology of Expectation*. The MIT Press.
- Jackendoff, R. (2009). Parallels and Nonparallels between Language and Music. *Music Perception*, 26(3):195–204.
- Meyer, L. B. (1956). *Emotion and meaning in music*. Univ. of Chicago Press, Chicago. Ill., paperback ed., [nachdr.] edition.
- Moumdjian, L., Moens, B., Maes, P.-J., Van Nieuwenhoven, J., Van Wijmeersch, B., Leman, M., and Feys, P. (2019). Walking to Music and Metronome at Various Tempi in Persons With Multiple Sclerosis: A Basis for Rehabilitation. *Neurorehabilitation and Neural Repair*, 33(6):464–475. Publisher: SAGE Publications Inc STM.
- Murata, H., Bouzarte, Y., Kanebako, J., and Minamizawa, K. (2017). Walk-In Music: Walking Experience with Synchronized Music and Its Effect of Pseudo-gravity. In *Adjunct Publication of the 30th Annual ACM Symposium on User Interface Software and Technology*, pages 177–179, Québec City QC Canada. ACM.
- Ready, E. A., McGarry, L. M., Rinchon, C., Holmes, J. D., and Grahn, J. A. (2019). Beat perception ability and instructions to synchronize influence gait when walking to music-based auditory cues. *Gait & Posture*, 68:555–561.
- Repp, B. H. and Su, Y.-H. (2013). Sensorimotor synchronization: A review of recent research (2006–2012). *Psychonomic Bulletin & Review*, 20(3):403–452.
- Schlenker, P. (2017). Outline of Music Semantics. *Music Perception*, 35(1):3–37.

- Schlenker, P. (2019). Prolegomena to Music Semantics. *Review of Philosophy and Psychology*, 10(1):35–111.
- Schubert, E. (2004). Modeling Perceived Emotion With Continuous Musical Features. *Music Perception*, 21(4):561–585.
- Styns, F., van Noorden, L., Moelants, D., and Leman, M. (2007). Walking on music. *Human Movement Science*, 26(5):769–785.
- Tieu, L., Schlenker, P., and Chemla, E. (2019). Linguistic inferences without words. *Proceedings of the National Academy of Sciences*, 116(20):9796–9801.
- Washburne, C. and Derno, M. (2004). *Bad music: the music we love to hate*. Routledge, New York. OCLC: 55981688.
- Willats, J. (1997). *Art and representation: new principles in the analysis of pictures*. Princeton University Press, Princeton, N.J.
- Zaradzki, L. (2021). *Les événements en sémantique linguistique et musicale*. PhD dissertation, Université Paris Cité.

Publications

Peer-reviewed articles

MIGOTTI, L. & GUERRINI, J. (2023). Linguistic inferences from pro-speech music: Musical gestures generate presuppositions, scalar implicatures, supplements and homogeneity inferences. *Linguistics and Philosophy*.

<https://doi.org/10.1007/s10988-022-09376-9>

Articles to be submitted

MIGOTTI, L., DECULTOT, Q., GRAILHE, P. and AUCOUTURIER, J-J. (2023). Cracking the pitch-code of motor-music synchronization using data-driven methods.

<https://psyarxiv.com/zkbn3/>

MIGOTTI, L., CHEMLA, E., SCHLENKER, P. (2023). How music can refer to objects: Coreference from dynamics, pitch and timbre in audiovisual animations.

MIGOTTI, L. & ZARADZKI, L. (2023). How music evokes walking: A case-study in music semantics.

Conference Proceedings

MIGOTTI, L. & ZARADZKI, L. (2019). Walk-denoting music: Refining music semantics. *Proceedings of the 22nd Amsterdam Colloquium*.

https://archive.illc.uva.nl/AC/AC2019/uploaded_files/inlineitem/1AC2019_Proceedings.pdf

ABSTRACT

Does music have meaning? And if so, what kind of meaning? In this dissertation, we provide experimental data supporting the view that music has a semantics: (i) music is able to refer to an extra-musical reality, and (ii) it does so in a rule-governed fashion. We present four experimental paradigms designed to test these two assumptions by having music interact with other cognitive faculties such as language, visual perception, or body motion.

In the first experiment, we presented participants with hybrid sentences containing music in lieu of words and tried to determine what meaning they get from such sentences. The results show that the informational content of embedded musical stimuli could be divided among the different slots of the linguistic typology (i.e. the different types of inference that complex sentences trigger). This suggests that the cognitive mechanisms classifying meaning in language can accommodate musical information.

In the second experiment, we had participants judge audiovisual scenes in which musical properties such as pitch height, dynamics, and timbre were manipulated and paired with visual objects. The results show that (i) both pitch height and dynamics can be semantically interpreted as the energy level or the distance of a visual object, (ii) pitch height and dynamics can be used to retrieve which object is being referred to when multiple objects are present, and (iii) timbre is interpreted as the very nature of an object, hence the timbre/object association cannot be violated in audiovisual scenes. Together, these results show that musical properties are not only associated with general scenes, but that they are bound to objects in those scenes.

In the third experiment, we built a case study about music that evokes walking. We sought to determine what it takes for a musical piece to evoke a walk through a preference task in which participants had to pick stimuli based on how well they evoked a character walking. The results show that at least five properties are involved in the association of a musical stimulus with a walk: three of which (steadiness, alternation, binarity) relate to structural properties of a walk, and two of which (absolute consonance and relative chord proximity) relate to physical properties of a walk such as stability. Consonance was arguably interpreted as the stability of each step, while chord proximity was associated with gait symmetry. The results further show that this fine-grained model accounts for the semantics of music evoking walks better than previous models and provide preliminary evidence that each musical event might have to be interpreted as a real-world event.

In the fourth experiment, we finally checked whether it is possible to find footprints of the semantics of pitch on body motion. Participants were tasked to walk in synchrony with a series of sounds varied in pitch. We found that their steps were both longer and produced with greater force on lower pitches. The similarity between the physical effect produced by lower pitch (mimicking an increase in weight) and the semantics of lower pitch (associated with bigger objects, or less energy) suggests that there is an interaction between musical representations and the motor system.

Together, our results provide evidence for a music semantics: music can refer to an extra-musical reality according to a set of systematic rules. They also show that music interacts with other cognitive systems either directly bearing meaning (language) or not (body motion), which raises questions as to how music-induced effects relate to musical meaning.

KEYWORDS

music, semantics, meaning, music cognition

