Holistic identification of musical harmony

A theoretical and empirical study

Ville Langfeldt

NMH Publications 2022:7

Norges musikkhøgskole
Norwegian Academy of Music

CERM
Centre for Educational Research in Music

Holistic identification of musical harmony

A theoretical and empirical study

Ville Langfeldt

Dissertation for the PhD degree
Norwegian Academy of Music, Oslo, 2022
NMH Publications 2022:7

NMH Publications 2022:7

ISSN 0333-3760
ISSN 2535-373X (online)
ISBN 978-82-7853-312-3 (print)
ISBN 978-82-7853-313-0 (online)

Norwegian Academy of Music
PO Box 5190 Majorstua
0302 OSLO
Tel.: +47 23 36 70 00
E-mail: post@nmh.no
nmh.no

Setting and printing: Bodoni, Oslo, 2022

Acknowledgments

This dissertation was written during years of great change and uncertainty in the world. Hopes for a better tomorrow have repeatedly been thwarted by a turn for the even worse. Through political and social unrest, climate crisis, a global pandemic, and international war, having the chance to immerse myself in the trivial subject of harmonic listening has been both absurd and a privilege—and a necessary escapism. For this possibility, I thank the PhD Committee at the Norwegian Academy of Music, and in particular Solveig Christensen for making it such a good experience.

I could not have finished this work without the help of others. I am especially grateful to my supervisor, Rolf Inge Godøy, for his wisdom and for being so committed to my project—even into retirement—and to my co-supervisor, Siw Graabræk Nielsen, for her steady hand in pedagogical and statistical matters.

A special thanks to Lawrence Zbikowski, Erkki Huovinen, Lasse Thoresen, Ivan Jimenez, and Tim Chenette, all of whom have offered valuable perspectives to my project, and to my music theory colleagues at the Academy—Kjell Habbestad, Anders Tykesson, Peter van Tour, Ram Reuven, Haakon Støring, and Bjørnar Utne-Reitan—for discussing harmony with me and sharing their insight.

I would also like to thank office mates Stine Camilla Blichfeldt-Ærø and Sunniva Thomassen for keeping me sane, Lise Hvoslef for walking me, Jakop Janssønn for making music with me, my students for asking questions I have to find answers for, the librarians at NMH—unsung heroes of academic life—for their tireless service, and Laura Macy, Tore Simonsen and Matej Marek, who have all done their utmost to amend the shortcomings of the written dissertation.

Finally, I would like to thank my parents and my sisters for their love, support, and patience with me during this highly selfish period of my life.

Oslo, May 2022
Ville Langfeldt

Summary

This dissertation explores *holistic harmony identification*. Aural training theorist Gary S. Karpinski depicts this as the most effective and precise form of harmonic listening, but a clear definition and understanding of its nature have been lacking. There has also been a lack of pedagogical strategies for advancing the development of such listening skills in aural training students—indeed, Karpinski questions whether such strategies are possible at all.

The dissertation consists of three parts. The first part is a theoretical study, in which the concept of holistic harmony identification is explored broadly through a combination of Gestalt theory and an ecological approach to perception. The study closes with a discussion of possible pedagogical approaches, including a "metaphorical" approach to harmonic listening. This approach is based on the claim that hearing a chord or progression *as* something is a way of acknowledging and verbalizing its holistic quality.

The dissertation's second part is a qualitative document analysis that further explores "metaphorical listening." Through an analysis of 20 textbooks on harmony, examples of cross-domain mapping in esthesic descriptions of chords and progressions are recorded and discussed. The main aim is to examine whether certain metaphor structures are more recurrent than others, which might suggest a relevance for harmonic aural training. The study's theoretical framework is conceptual metaphor theory.

The third part is a statistical study that examines one of the metaphors found in the document analysis: *harmonic luminosity*, or the idea that harmony can express "brightness" and "darkness." The phenomenon is examined empirically through a web-based experiment with 236 participants. The study shows that harmonic luminosity is likely more perceptually complex than it is portrayed in textbooks using this metaphor. The results also indicate that harmonic luminosity might require some degree of perceptual learning.

The dissertation's main contribution is a conceptual framework for holistic harmony identification, which both elucidates the concept and enables more targeted pedagogical approaches. In a further exploration of such approaches, novel perspectives on the role of metaphor in musical harmony are offered.

Sammendrag

Denne avhandlingen utforsker *holistisk harmonisk identifikasjon*. Evnen til slik lytting er av gehørteoretiker Gary S. Karpinski beskrevet som den mest effektive og presise formen for harmonisk gehør, men har hittil manglet en klar definisjon. Det har også manglet didaktiske strategier for å fremme slike evner hos gehørstudenter, samt enighet rundt hvorvidt didaktiske strategier overhodet er mulige.

Avhandlingen består av tre deler. Første del er en teoretisk studie, hvor holistisk harmonisk identifikasjon undersøkes bredt ved hjelp av gestaltteoretiske prinsipper og en økologisk tilnærming til persepsjon. Denne studien avsluttes med en diskusjon av mulige didaktiske tilnærminger, inkludert en «metaforisk» tilnærming til harmonisk lytting. Denne tilnærmingen baserer seg på påstanden om at det å høre en akkord eller akkordprogresjon *som noe* innebærer å anerkjenne og sette begrep på dens holistiske kvalitet.

Avhandlingens andre del er en kvalitativ dokumentanalyse som undersøker «metaforisk lytting» nærmere. Gjennom analyser av 20 engelskspråklige lærebøker som omhandler harmonikk, kartlegges metaforer som brukes i beskrivelsen av akkorders og akkordprogresjoners klanglige fremtoning. Målet er å undersøke hvorvidt bestemte metaforstrukturer har en mer allmenn utbredelse enn andre, og dermed kan ha relevans for gehørundervisningen. Det teoretiske grunnlaget for analysen er kognitiv metaforteori.

Tredje del er en statistisk studie som omhandler én av metaforene som ble kartlagt i dokumentanalysen: *harmonisk luminositet*, eller tanken om at harmonikk kan uttrykke «lys» og «mørke». Fenomenet undersøkes empirisk gjennom et nettbasert eksperiment med 236 deltakere. Studien viser at harmonisk luminositet sannsynligvis er mer perseptuelt komplekst enn det gis inntrykk av i lærebøkene som benytter metaforen. Resultatene antyder også at harmonisk luminositet forutsetter en grad av perseptuell læring.

Avhandlingens primære bidrag er et konseptuelt rammeverk for holistisk harmonisk identifikasjon, som både klargjør konseptet og åpner for mer målrettede didaktiske strategier. Gjennom en videre utforsking av slike strategier bidrar avhandlingen også med nye perspektiver på metaforers betydning i musikalsk harmonikk.

Contents

Figures

PART I

On Gary S. Karpinski's *Gestalt listening:* A theoretical exploration and conceptual framework

1 Introduction

1.1 The paradox of Gestalt listening

In a short—and, at first sight, rather inconspicuous—passage in his landmark book *Aural skills acquisition: The development of listening, reading, and performing skills in college-level musicians* (2000), Professor Gary S. Karpinski describes something he calls "Gestalt listening." In this section of the book, Karpinski lays out commonly used listening approaches in harmonic dictation exercises.[1] Among well-known strategies such as part writing, arpeggiation, listening for the bass line, and so forth, he mentions an approach that is somewhat more mysterious and elusive. Since the passage sparked an interest in me that eventually grew into this dissertation, I will quote it in almost its entirety. (The heading, not included here, is *Gestalt*.)

> In yet another approach, listeners are asked to learn to identify chords as complete entities—to recognize the "subdominantness" of the subdominant chord, for instance. This goal seems reasonable enough: an integral aspect of many expert listeners' strategies involves a certain amount of raw, whole-harmony recognition. But whereas such Gestalt listening is in the end quite rapid and fluid, it is difficult to develop directly. Listeners at any given stage of development either recognize a particular chord or don't. In the absence of other listening strategies, they can be taught no concrete means for recognizing a "new" chord as a Gestalt. Instead, we might think of Gestalt listening as a by-product or result of other techniques. After weeks, months, or even years of repeatedly recognizing and labeling particular chords, those chords can become instantly recognizable—in the same Gestalt manner we recognize a well-known face, a familiar voice on the telephone, or the taste of a common spice. Over time, individual chords, harmonic groups—indeed, many musical figures—drop out of the class of stimuli that require intellectual scrutiny in order to be perceived and take a place in a personal pantheon of essentially instantaneously recognizable entities. (Karpinski, 2000, p. 119)

There are several interesting points to consider in this extract. One is the fact that "subdominant" is a harmonic function dependent on a tonal context, and not a specific type of chord structure. Thus, when Karpinski talks about recognizing "particular chords," he hints at something that is not pursued further: There must be at least three aspects to harmonic

1 Karpinski's aim is to "[examine] a few of the more commonly taught methods" (Karpinski, 2000, p. 118). In other words, his list of harmonic dictation approaches is descriptive, not necessarily normative.

Gestalt listening. On the one hand there is the intrinsic, static quality of a specific structure of pitches, i.e., the *single chord* as a Gestalt. The major triad will always form a Gestalt which differs from a minor triad. On the other hand, single chords are influenced by the harmonic context in which they appear. Thus, Karpinski's subdominant example also has a temporal or contextual dimension. In addition to these two aspects of the single chord, Karpinski mentions that *harmonic groups* may also be identified as a Gestalt. This must rely on sequential grouping (i.e., hearing a group of chords as one Gestalt, such as "descending fifths sequence"), which seems to be yet a different aspect of Gestalt listening. For now, I will simply point out this complexity, since Karpinski does not do so himself. Later in the study, however, these distinctions will be thoroughly discussed.

Most striking to me, however, is the pedagogical paradox embedded in Karpinski's concept: Gestalt listening is an *expert's* way of hearing harmony, both *rapid and fluid*. Identification comes *instantaneously* and demands *no intellectual scrutiny*. Even if one has not experienced this kind of familiarity in the area of harmonic listening oneself, one might be able to relate to Karpinski's examples of tasting a familiar spice or hearing a family member's voice on the telephone. So how does one go about acquiring this kind of aural skill? Aye, there's the rub: As it turns out, Gestalt listening is difficult to develop directly, according to Karpinski. There are no listening strategies to be taught. One may only hope that perhaps, after years of immersing oneself in harmony, one suddenly finds oneself able to recognize the "subdominantness" of a subdominant chord (or the "omnibus-ness" of an omnibus progression,[2] to use a more complex harmonic example) with the same effortlessness that one recognizes one's mother's voice.

But is it necessarily so? The shrug with which Karpinski rejects the possibility of targeted Gestalt listening training, stands in strange contrast to the potentially considerable pedagogical gain of such training methods. Karpinski is hardly to blame: His book is not about Gestalt listening. Nor should he be expected to elaborate on a concept he references primarily because it is a relatively "commonly taught method" (Karpinski, 2000, p. 118), but which he himself raises pedagogical concerns about. However, it would seem that before we conclude on this matter, we must at the very least have a qualified understanding of what Gestalt listening *is*. Karpinski's description is rather superficial and external, emphasizing the *outcome* of such listening (i.e., instant identification) rather than, for example, its genesis, its perceptual nature, or the relationship between perceptual stimulus and the perceiver. Such questions are the subject of the present dissertation.

2 Omnibus: "A type of prolongation (usually of the dominant function) where one or more upper voices move chromatically and in contrary motion to the bass, which ascends or descends chromatically. The result is a symmetrical division of the octave by minor thirds" (Laitz, 2016, p. G-11).

1.2 Why Gestalt listening?

Am I reading too much into the term "Gestalt listening," briefly mentioned, almost in passing, by a single theorist? Perhaps Karpinski coined it on a whim, as an umbrella term for a range of perceptual phenomena? That may be so. But Karpinski is not the only one in the music theory literature who describes a non-analytical and "holistic" type of harmonic recognition (i.e., recognizing the "overall sound" of chords). On the contrary, versions of it are readily found in texts dealing with harmony perception. However, only a few authors give it even a rudimentary description (e.g., Rahn & McKay, 1988)—more often it is mentioned rather casually, without much context or instructions (e.g., Coker et al., 1997, p. 35; Gorow, 2006, p. 259; Laitz, 2008, p. 253; Radley, 2008, p. 87; Thackray, 1993, p. 125).

From the perspective of aural training teaching methods, this "expert skill" is worthy of attention no matter what one would call it. It is neither restricted nor fully explained by a single label. I have singled out Karpinski for a couple of reasons: First, he gives the phenomenon a name. While he might perhaps as well have chosen a different label altogether, I believe that the word *Gestalt* encapsules an important aspect of the phenomenon. It signifies something that is grasped immediately and spontaneously, seemingly without conscious perceptual analysis. The descriptive power of this term should not be undervalued. As noted by Leman and Schneider (1997), Gestalt concepts have "often been acknowledged for their plausibility in terms of human experience, and thus, closeness to the *real world*" (p. 17). Second, Karpinski holds up Gestalt listening as a kind of gold standard for harmonic listening skills. It is not merely one listening strategy out of many, it is the *end goal*, something one can only hope to obtain after years of experience—according to Karpinski. Something as desirable as this must surely be examined closely, so that we can both understand better its perceptual and phenomenological nature and better assist students in developing such skills for themselves.

1.3 Why this study?

The topic of harmony perception has long been somewhat neglected in aural training literature. As an example, a keyword search for "aural skills" in SAGE Journals (*Psychology of Music*, *International Journal of Music Education*, *Journal of Research in Music Education*, *Music Educators Journal*, *Research Studies in Music Education*, *Update: Applications of Research in Music Education*, *Journal of Music Teacher Education*, *General Music Today*) produces 39 music-related articles published in the period 2004–2021.[3] Of these 39, only two are concerned

3 The search was not restricted to this period.

with harmony (Jimenez & Kuusi, 2018a; Woody, 2020). In comparison, 20 deal with aspects of melody (i.e., melodic dictation, interval training, sight-singing, and melody memorization). While this cross section of studies naturally does not constitute the entire body of relevant research, it is indicative of a general bias in the focus of aural training: Melody is given much more attention than harmony.[4] A similar tendency can be found in the pedagogical literature. In his book *Aural awareness: Principles and practice*, George Pratt (1998)—with Michael Henson and Simon Cargill—sets out to amend the bias towards melody and rhythm in traditional aural training, by giving attention to other, neglected dimensions of music. In the second chapter, he presents "the elements of musical expression": Meter; rhythm; pitch; texture; timbre; compass, range, and density; dynamics; articulation; placing in space; pace; and structure—but not harmony.

Nevertheless, the picture is not as bleak as it may sound. Most traditional aural training books have at least one chapter on harmony, and there are even books entirely dedicated to harmonic hearing (e.g., Coker et al., 1997; Radley, 2008). Even so, Karpinski's concept of "Gestalt listening" represents a scarce resource in this landscape: a portrayal of an *expert's* way of listening. Harmonic ear training literature is often more concerned with "practical tools" for the uninitiated student (cf. Johnson, 2013) than with the skills of the virtuoso. Furthermore, Karpinski indicates that this expert skill operates in a fundamentally different way than the atomistic, constructional way of harmonic listening which students are traditionally subjected to in aural training (i.e., the other strategies mentioned by Karpinski, such as part writing, arpeggiation, listening for the bass line, etc.). While the latter approach is pedagogically sensible—you must learn to crawl before you can walk—this discrepancy raises the question of whether there are other pedagogical approaches which are better aligned with a "Gestalt" way of listening. Karpinski, as shown, suggests there are not.

However, if we are to accept "Gestalt listening" as the ultimate goal for harmonic ear training, we in the aural training community owe it to ourselves—and to our students—to shed some light on its nature. We should be interested in knowing more about the ways of listening that may fit this description. We should want to form some hypotheses about how such listening skills come about, and what characterizes a Gestalt identification of harmony. And not least, we should attain a qualified opinion on whether it is learnable or teachable. This means exploring whether there could be ways of teaching harmonic listening in a truly holistic manner, rather than as melody perception multiplied.

4 This imbalance is almost exactly mirrored by the heavy bias for harmony in music theory.

1.4 Previous research

There is limited research on the concept of "Gestalt listening" as such (i.e., with reference to Karpinski). Jarvis (2015) discusses pedagogical approaches to developing Gestalt listening skills through maximizing students' exposure to both challenging single chords and common harmonic progressions. His approach is based on statistical learning, i.e., presenting students with "enough repetitions, reevaluations, and varying conditions that will allow them to dramatically increase their ability to hear harmony holistically" (Jarvis, 2015, para. 13). In essence, Jarvis's approach adheres to Karpinski's claim that Gestalt listening skills can only evolve through "weeks, months, or even years of repeatedly recognizing and labeling particular chords" (Karpinski, 2000, p. 119). It also faces the same challenge: In the absence of more specific holistic listening strategies, there is no guarantee that extensive repetition in itself will produce the desired result. The statistical learning approach is based on the assumption that "brains are sorting through the continuous data and strengthening correct neurological connections or reevaluating incorrect ones" (Jarvis, 2015, para. 10).

In an observational study titled *Does Gestalt hearing exist?*, Chenette et al. (2021) explore the harmonic listening strategies reported by 73 participants. In addition to completing a harmonic dictation test, the participants reflected, in their own words, on how they typically identify a chord progression. Subsequently, they were asked to identify their own listening strategies from a pre-made list. Included in this list were the options "Expectation of what is likely to come next," "Chord quality (major, minor, etc.)," "Bass line," "The feeling of a chord (the 'one-ness' or 'it feels dominant-y')," "Solfege syllables or scale degrees in the chords," "Cadences," "Imagining playing the chord on some instruments," and "Other" (Chenette et al., 2021, p. 150) While yielding interesting results, the study illustrates the difficulty in examining Karpinski's concept without a clear definition. For example, is Gestalt hearing—represented by the "the feeling of a chord" option—sharply distinguishable from other parameters presented by Chenette and colleagues, such as hearing the bass or chord quality, or predicting what chord will come? It seems that in Gestalt identification of a chord or progression, the way Karpinski portrays it, more than one of these categories must be active simultaneously. If knowledge of bass tone and chord quality are not included in a Gestalt identification, then what is? This raises some questions about the methodological approach: First, is identifying the bass line—the strategy most frequently reported by the participants who scored 100 percent on the harmonic dictation—by necessity incompatible with (or separable from) Gestalt hearing? This depends, in part, on what exactly is meant by "Gestalt hearing."

Broadening our scope from the specific terms "Gestalt listening" or "Gestalt hearing," a multitude of research and literature carry a connection to the subject of the present theoretical

study. This includes research on harmony perception (e.g., Jimenez & Kuusi, 2018a; Jimenez & Kuusi, 2018b; Jimenez, Kuusi, Czedik-Eysenberg, et al., 2021; Jimenez et al., 2020; Jimenez, Kuusi, & Ojala, 2021; Kuusi, 2013), research on holistic perception (e.g., Bregman, 1990), musicological contributions (e.g., Doll, 2017; Gjerdingen, 2007), and theoretical contributions in aural training literature (e.g., Peebles, 2021; Rahn & McKay, 1988; Stevens, 2016, 2020). Because of its scope and contextual relevance, appropriate parts of this research will be presented and discussed at the relevant places later in this study.

1.5 Research questions for the theoretical study

I do not believe that it is possible to fully explain "Gestalt listening" once and for all—mainly because it is not necessarily *one thing*, ontologically speaking. Rather, I believe it is probably best viewed as a collective term for "holistic" identification of harmony, involving a range of factors such as memory, knowledge, perceptual learning, enculturation, association, embodiment, multimodality, emotion, social context, etc. In a strict scientific sense, then, the question "What is Gestalt listening?" may not be answerable at all. I do not, however, find that this is cause for resignation. On the contrary, I will spend the first part of this dissertation trying to find a theoretical perspective that may unify—or at least encompass—all these components and provide us with a functional interpretation of the term. What we *can* shed some light on, through a theoretical exploration of the phenomenon, is its more pedagogically relevant aspects. Therefore, the first third of this dissertation is a theoretical study of the phenomenon of "Gestalt listening"—or, as I will eventually propose to rename it: *holistic harmony identification*.[5]

The theoretical study will concentrate on three research questions:

1. **What does it *mean* to identify harmony holistically?** This question addresses the *concept* of holistic harmony identification, seeking a usable definition that delimits it from other types of harmonic listening. I will discuss the terms "holistic," "harmony," and "identification" independently, using literature on aural training, knowledge, learning, and music perception, in order to elucidate the meaning of their combination.

2. **What *characterizes* holistic harmony identification?** This question addresses the act of holistic harmony identification and the perceptual environment it takes place in. Using a predominantly ecological approach (cf. Gibson, 1966, 1979), I will explore various aspects of harmonic listening in the aural training classroom, including listener

5 The argumentation for this choice is found under Chapter 2.4.3.

intentions, perceptual meaning, perceptual learning, and harmonic invariants. The pedagogical relevance of these will be discussed.

3. **Can holistic harmony identification be taught?** This question addresses Karpinski's pedagogical paradox and reassesses the claim that holistic harmony identification cannot be trained deliberately. I investigate some possible trajectories for such training, based on the answers to RQ2.

Before turning to these questions, however, I will present and discuss a theoretical framework upon which this study will be founded.

2 Gestalt theory

Regardless of how Karpinski's term *Gestalt listening* will be treated later in this dissertation, I find it necessary to start with a brief introduction to Gestalt theory. Not because I think it is necessarily the best theoretical prism for examining the phenomenon—as will eventually become clear—but because the term *Gestalt* itself is so deeply entangled with the history and diversity of the Gestalt psychology movement. Irrespective of how the term might be tossed around today, it carries within it a rich scientific tradition and a distinct view on perception. To ensure this context is not ignored, I will give a short overview of Gestalt theory's history, its central concepts, and how it has influenced music research.

2.1 Historical outline and key concepts

While building on ideas and concepts already put forward by others, Gestalt psychology as a distinct movement of thought is generally considered to have started with Max Wertheimer's (1912) paper on a visual illusion called apparent motion, or phi-motion.[6] The Gestaltists' main tenet was that percepts are organized into structured wholes, or *Gestalten*, which are the primary units of mental life (Wagemans, 2015, p. 5). This view was a rejection of elementarism, the belief that perceptions are best described by examining the elementary parts of experience (Rock & Palmer, 1990, p. 84).

Though often associated with the study of vision, the Gestalt psychology movement had musicological roots. The concept of *Gestalt* itself can be traced back to Christian von Ehrenfels's theoretical work in the late nineteenth century. Ehrenfels was a philosopher, but he had also received lessons in composition from Anton Bruckner and written librettos for a number of music-dramas (Smith, 1988, p. 12). In a seminal essay titled "On Gestalt Qualities" ("Über 'Gestaltqualiteten'"[7]) (1890/1988), he seeks to answer a seemingly simple question: What is a melody? He infers that it is *something more* than the sum of the individual tones it contains. Otherwise, how would we still recognize it as the same melody after transposing it to a different key, thereby replacing all its components? (1890/1988, p. 90) This "something more" is what Ehrenfels calls a *Gestalt quality*. Listening to a melody, we attribute to it a sense of organized

6 Apparent motion is the visual effect that occurs when, for instance, lightbulbs around an old-fashioned billboard are switched on and off in specific patterns, creating the illusion of movement. It is also the perceptual principle behind film, where a number of images (frames) are shown at a such a high rate that it is perceived as a *motion picture*. Apparent motion has later been seen as a visual parallel to the phenomenon of auditory streaming (Bregman, 1990, pp. 17ff).

7 The German word and concept *Gestalt* has no straightforward English translation. This is the reason it has been adopted into English (and other) languages.

whole, which is based on its individual tones but is at the same time more than these elements. That this Gestalt quality is a projection of the listener's mind, and not an attribute of the tone series itself, is demonstrated by the fact that "at most *one* element [i.e., one melody tone] can be given in perceptual presentation, the remainder being present via memory-images (or images in expectation relating to the future)" (1890/1988, p. 94).

Another important forerunner was philosopher and psychologist Carl Stumpf, who, in his major work *Tonpsychologie* (1883, 1890), describes the phenomenon of *Verschmelzung*, or "tonal fusion." This is an effect whereby harmonic intervals, under certain circumstances, are prone to "melting together" perceptually. Stumpf attributed this effect both to structures inherent in the percept, and, significantly, to *cognitive* function—in other words, *Verschmelzung* is *triggered* in the subject's perceptual process by certain formations (Schneider, 1997, pp. 118–119).[8] Stump was also the teacher of Gestalt psychologists Wolfgang Köhler and Kurt Koffka (and, incidentally, of Edmund Husserl).

Wertheimer's 1912 paper marked the start of an influential movement in psychology, dedicated to theoretical and empirical research on perception, learning, memory, reasoning, social relations, and aesthetics. Three main schools can be distinguished: the Berlin school, to which Wertheimer belonged, the Leipzig school, and the Graz school. Although they had their disagreements, for example regarding the relation between the mental and the physical (Eichert et al., 1997, p. 71), their differences were outweighed by their shared rejection of the dominant approach in perceptual psychology (Tan, 2015, p. 109). One of the most prominent contributions was the description of laws of perceptual grouping—commonly referred to as "Gestalt laws"—formulated by Wertheimer (1923/1938). These state that elements are prone

8 Stumpf's concept of *Verschmelzung* has been understood (and perhaps misunderstood) in different ways. I will not pursue this in detail, as it is of little consequence for the present context, but merely sketch the outline of the debate: The psychoacoustic tradition defines tonal fusion as "fooling the scene-analysis process" (Bregman, 1990, p. 508), whereby the harmonics of two tones blend to a degree where they may actually be perceived as one tone. For this reason, the phenomenon is also referred to as *harmonic* fusion rather than *tonal* fusion in more recent literature (Huron, 2016, p. 224). In this sense, chords are unlikely to be susceptible to tonal fusion, as only a lack of musical competence will cause mistaking three or more tones for one. Furthermore, from this interpretation of *Verschmelzung*, Stumpf was clearly wrong to postulate it as the cause of consonance (Huron, 2016, p. 59). On the other hand, Schneider (1997) argues that *Verschmelzung* is *not* to be understood in such a way that the individual tones in an interval or triad are no longer *discernible* to the listener. On the contrary, "if we are listening to a completely harmonious chord, we can focus on one or several constituents (notes and/or their respective partials) for a moment, then return to the chord as a complex whole, then return to another constituent realizing its relation to other parts as well as to the overall configuration, etc. Thereby, our attention kind of *swings* between parts and whole, however, we are attentive to but one stimulus configuration that is *looked at* (listened to) from different angles" (Schneider, 1997, p. 121). This is clearly a precursor for Gestalt psychology. Bregman blames the confusion over *Verschmelzung* on Stumpf, claiming he "failed to properly distinguish between 'heard as one' and 'heard as smooth'" (1990, p. 508). However, Stumpf did unmistakably state that "Uniformity is not unity. *Verschmelzung* does not mean that one is completely unable to tell the two tones apart," ("Einheitlichkeit ist nicht Einheit. Verschmelzung bedeutet nicht, dass man die beiden Töne überhaupt nicht unterscheide" *my translation*] (Stumpf, 1911, p. 118).

to perceptual grouping when close together (i.e., *the law of proximity*), when similar to one another (*the law of similarity*), when forming a closed shape (*the law of closure*), or when perceived to move in the same direction (*the law of good continuation*) (Rock & Palmer, 1990, p. 85). Wertheimer did not view such grouping as a sort of active "strategy" on the perceiver's behalf, but as an inherent and mostly subconscious tendency in our perception. Although he presented his Gestalt laws with examples predominantly from visual perception, they were considered valid for all sensory modalities, including sound and music.[9] Illustrative of the lineage from Ehrenfels's early philosophical reflections on melody to modern music psychology, most present day models of melodic perception acknowledge the perceptual laws of Gestalt theory (Radocy & Boyle, 2012, p. 292).

During the 1930s, the Gestalt movement was diminished due to several circumstances: The rise of Nazism in Germany and the subsequent emigration of central figures to the United States; institutional affairs; a paradigmatic shift toward behavioristic psychology; and a lack of clarification concerning conceptual and ontological issues (Leman & Schneider, 1997, p. 15; Smith, 1988, p. 70; Stadler, 1985, p. 139; Wagemans, 2015, p. 15). However, in the course of the same decade, Gestalt psychology was accepted into American mainstream psychology (Luchins, 1975, p. 32).

2.2 Present status and criticism

Today, the scientific status of Gestalt psychology is somewhat unclear. Many of the Gestaltists' theories about brain function have been refuted by modern neurophysiology (Rock & Palmer, 1990, p. 90). Another common criticism, raised already in 1949 by Hebb (referenced by Meyer, 1956, p. 84), is that Gestalt theory's focus on spontaneous organization leaves little room for experience and perceptual learning. These and other criticisms were moderated in a large historical review of the movement's first century (Wagemans et al., 2012), but Gestalt psychology remains controversial. Nonetheless, Gestalt psychology exerted influence on both general psychology (Leman & Schneider, 1997, p. 16) and on cognitive psychology (Flores d'Arcais, 1975; Rock & Palmer, 1990) throughout the second half of the twentieth century.

Moreover, Wertheimer's laws of perceptual grouping are still central to the discourse in the field of perception, and they have formed the basis for a substantial amount of modern research on perceptual grouping (Brooks, 2015, p. 62). Most contemporary psychology textbooks include a Gestalt-like chapter, and empirical research is published regularly (Wagemans, 2015, p. 3).

9 For a discussion of the Gestalt laws in terms of sound, see Denham and Winkler (2015).

Criticism of poorly defined concepts—a weakness already pointed out by Wertheimer himself (1924/1938, p. 3)—has been met with conceptual clarification, creating a modern notion of Gestalt perception supported by fundamental neurophysiological data (Guberman, 2017).

Not least, Gestalt thinking has had a great and lasting impact on music psychology and research on music, something I will cover in more detail.

2.3 Applications of Gestalt principles to music research

2.3.1 Gestalt and music perception

There is a long, broad, and persistent tradition for using Gestalt principles—implicitly or explicitly—to explain different aspects of music perception. I will mention just a few examples.

Already parallel to the emergence of the Gestalt movement, *musical hearing* ("musikalisches Hören") became a key word in German music literature (Hovland, 2009). The concept emphasized the active participation of the musical mind in structuring and transforming "arbitrary acoustic events … into a coherent whole" (p. 180). This creative aspect of music listening is not least emphatically present in Ernst Kurth's influential *Musikpsychologie* (1931), which explicitly looks to Gestalt psychology (Tan, 2015, p. 123). In a similar vein, Leonard B. Meyer uses Gestalt principles to explain the construction of musical meaning in his classic, *Emotion and Meaning in Music* (Meyer, 1956).[10]

In modern mainstream music psychology, Diana Deutsch (e.g., 1982, 2013) has been a central figure in using Gestalt theory to explain the perceptual organization of music. From a more compositional/aesthetic standpoint, Tenney (2000) examines how Gestalt factors of cohesion and segregation influence the formation of aural Gestalts ("clangs") in the perception of twentieth century music, and how these are then phenomenologically grouped into larger musical gestalts (sequences). Gestalt theory is also a significant influence in *aural sonology* (e.g., Thoresen, 2015), an analytical approach to music-as-heard from multiple phenomenological perspectives, developed by faculty members at the Norwegian Academy of Music.

Two especially influential works should be highlighted: Fred Lerdahl and Ray Jackendoff's book *A Generative Theory of Tonal Music* (1983) examines the "grammar" of music, using

10 Although Meyer makes a point of distinguishing between the speculative aspects of Gestalt theory itself, and its empirical findings (Meyer, 1956, p. 83).

on the one hand a linguistic approach, inspired by Chomsky's transformational-generative grammar, and on the other, a Gestalt approach. The work is both an examination of the structural properties found in traditional (mostly Western) music, and, more importantly, of the cognitive processes involved in "making sense" of these structures. The grouping preference rules presented in the book (derived from Wertheimer's Gestalt laws) have more recently gained some empirical support (Deliege, 1987), though they have also been criticized for vagueness (Hansen, 2010/11, p. 40). Lerdahl and Jackendoff's theory (commonly referred to as *GTTM*) shifted the general focus of music analysis from the traditional work-oriented approach, towards an increased interest in the listener's cognitive percept, thus introducing the "cognitive paradigm" (Hansen, 2010/11, p. 34).

A similarly seminal work, in the field of psychophysics, is Albert S. Bregman's *Auditory Scene Analysis* (1990). Bregman uses Gestalt laws to explain organizational processes in basic auditory perception, and indeed states that all Gestalt grouping principles apply to the field of hearing. For instance, when we detect sounds that strongly resemble each other (i.e., the law of *similarity*) or occur close in time (i.e., the law of *proximity*), we are likely to attribute them to the same source or happening (Bregman, 1990, p. 24). Such auditory grouping furthermore allows for the construction of hypotheses about the environment, for example expectations about what sounds might follow. This has led to a more modern view of auditory grouping principles as a heuristics acquired through evolution and learning, rather than as laws of physics, as was proposed by the Gestalt psychologists (Denham & Winkler, 2015, p. 603).

2.3.2 Gestalt and harmony

Although Gestalt approaches seem to have been applied mostly to broad examinations of grouping principles in general music perception, some individual studies have targeted more specific musical parameters, such as melody (Gjerdingen, 1994), *virtual* melody[11] (Schulte et al., 2001), rhythm (Fraisse, 1975; Klapp & Jagacinski, 2011; Klapp et al., 1998), music notation (Eden Ünlü & Ece, 2019), and even genre (Gjerdingen & Perrott, 2008). Harmony—the subject of the present study—has also been examined using a Gestalt perspective, although often in an implicit way which does not make active use of Gestalt theory. Rudimentary descriptions of what could, with a more general description, be called *holistic chord recognition* have been made by Rahn and McKay (1988), Karpinski (2000), and Tagg (2013), though with greater emphasis on pedagogical considerations than on the nature of the phenomenon itself.

11 *Virtual melody*: A "melody" consisting of non-sounding fundamental frequencies, indirectly implied through a (sounding) selection of their harmonics. In the study by Schulte et al., the melody *Frere Jacques* was implied through complex tones with missing fundamentals.

Widening the scope from single chords, Eberlein (1997) proposes a rough, empirically informed model for harmonic successions in tonal music through the recognition of harmonic-melodic stock patterns ("Gestalten"). The model involves comparing such intervallic features in sounding music to patterns already in the listener's cognitive repertoire, allowing for both identification and generalization. This is reminiscent of how Coker et al. (1997) present strategies for internalizing typical harmonic patterns found in jazz standards. In a similar way, although with greater emphasis on musical construction than on perception, Gjerdingen (2007) shows how the compositional "grammar" of the galant style consists of a stock repertoire of melodic-harmonic figures. While Gjerdingen calls his work a *schemata theory*, it has a clear Gestalt influence in how melody, harmony and formal placement combine into holistic "packet[s] of knowledge," which also constitute perceptual Gestalts for the experienced listener (Gjerdingen, 2007, p. 11). In a publication somewhat related to Gjerdingen's in methodology and purpose, Doll (2017) has more recently identified and characterized harmonic schemata in anglophone pop and rock music since the 1950s.

2.4 Gestalt concepts used in this dissertation

I do not believe that Gestalt theory alone is sufficient to examine Karpinski's concept of Gestalt listening—notwithstanding the label he gives it. Partly because Gestalt theory does a better job of describing perceptual organization than it does at explaining it. However, while many of its explanatory models—or lack thereof—remain contested (see, e.g., Bruce et al., 1996), Gestalt *principles* are seen by many as a fruitful heuristics for perceptual research (Bonacchi, 2015; Bregman, 1990; Denham & Winkler, 2015; Eichert et al., 1997; Leman & Schneider, 1997). Following this view, the present dissertation will adhere to Gestalt concepts only inasmuch as they provide useful conceptualizations of perceptual processes, rather than taking a dogmatic stance.

While Gestalt theory includes a range of concepts, I will primarily use two of them: Firstly, the term "Gestalt" itself (including "Gestalt quality," which refers to the perceptual quality present in a Gestalt). Secondly, the concept of *emergence*.[12] As will shortly become clear, these two terms are really two aspects of the same concept.

One might criticize me for largely ignoring the so-called "Gestalt laws," i.e., the principles that dictate perceptual grouping. These principles, as already mentioned, have had a great impact on

12 The concept of emergence is also used in philosophy and cognitive science (Stephan, 2006), and is not exclusive to Gestalt theory.

perception research in general and on music perception research especially, and they have not been replaced by later research (Leman & Schneider, 1997; Rock & Palmer, 1990)—although some consider them "descriptions of organizational tendencies" (Pomerantz & Kubovy, 1986, cited from Eichert et al., 1997, pp. 80–81) rather than actual "laws" of perception.

To address this criticism: My argument is *not* that the Gestalt laws of perceptual grouping are not relevant to the perception of harmony. On the contrary, to argue that they *do* apply to pitch configurations seems superfluous. It is established beyond reasonable doubt that listeners enculturated in Western harmony (at the very least) generally perceive configurations of pitches as auditory objects (i.e., "chords"[13]) rather than as individual, independent pitches (cf. Bregman, 1990, p. 496). While the perception of harmony is less straightforward in a historical perspective,[14] any modern textbook on Western harmony is a testament to the fact that chords are considered *distinct objects*, both perceptually and theoretically. They have names, symbols, and functions; they can be inverted or have tones added to them; they have neighbor, substitution, or resolution chords; and they combine to form progressions or sequences. And perhaps most telling: The constituent pitches of a given C major chord may be replaced entirely by other pitches (i.e., the tones C, E, and G in other octaves), yet one might be naturally inclined to call the resulting sonority "the same" as the first one. This demonstrates that harmonic Gestalts are more perceptually salient than the individual pitches from which they emerge. Clarke (2005) notes that this is both a reasonable and "correct" perception:

> (C)hords, and especially chords played on the piano, typically consist of notes with closely synchronized onsets, homogenous timbres, and very similar dynamic levels—all of which help to produce fusion between the chord components, as Bregman (1990, pp. 490–493) points out. So it is perfectly appropriate to hear a triadic chord as a single "thing": it *is* a single thing. (Clarke, 2005, p. 24)

Thus, rather than making a superfluous argument for the role of grouping principles in pitch simultaneity perception, I will focus on what *characterizes* the perception of such simultaneities. For this purpose, the concepts of *Gestalt* and *emergence* offer perspectives that I believe are useful. The terms, however, demand some closer discussion, so that it is clear what is understood by them.

13 Under certain conditions, to be clear. Not all pitch simultaneities will necessarily form structures that would normally be referred to as "chords" in the traditional sense of the term.

14 Cf. the historical European development from viewing sonorities as a non-conceptualized result of counterpoint and voice leading, to the modern notion of the "chord" as the basic unit of harmony (Cohn et al., 2001).

2.4.1 Gestalt and emergence

To recapitulate, we saw that Ehrenfels introduced the term "Gestalt" in 1890 to explain what a melody is. In a later text, he sums up his argument:

> The most obvious answer: the sum of the individual tones which make up the melody. But opposed to this is the fact that the same melody may be made up of quite different groups of tones, as happens when the same melody is transposed into different keys. If the melody were nothing other than the sum of the tones, then we would have to have here different melodies, since different groups of tones are involved. (Ehrenfels, 1932/1988, p. 121)

Thus, a melody is more than the sum of its tones, according to Ehrenfels. It has a certain quality that allows us to recognize it as "the same" melody even if all the individual tones are substituted—a *Gestalt quality*.

A Gestalt quality, as theorists would later elaborate on Ehrenfels's term, presupposes an underlying structure. However, not all structures possess a Gestalt quality. For a Gestalt quality to be brought into realization, the structure must possess a certain *Prägnanz*[15] (Metzger, 1965, p. 416). In somewhat simpler terms, *Prägnanz* means that the individual elements of the structure are part of a pattern that is "as simple, regular, balanced, or coherent as possible" (Metz-Göckel, 2015, p. 22). The parts will thus have a strong tendency towards perceptual grouping.[16] To expand on Ehrenfels's example of melodies, we could for instance compare a children's song to a dodecaphonic tone row. The former might typically use diatonicism, repetition, melodic sequencing, and smaller intervallic movement, all of which facilitate the perception of the song as a coherent and easily memorized melody—a *good form* (perceptually speaking, not esthetically). As for the latter, one could simply recall Anton Webern's optimistic comment that "sometime in the future even the postman will whistle my melodies!" (Moldenhauer & Moldenhauer, 1978, p. 543) A prophecy to which Steve Reich laconically remarked, more than half a century later, "there is no postman on earth who whistles his tunes" (Fulker, 2011).[17]

---

15 Metzger—consciously or not—creates the neologism "pregnance" when trying to anglicize the German word *Prägnanz*, while others state that the term cannot be translated (Wertheimer, 1923/1938, p. 79, footnote). Gurwitsch (1979) uses the term "perceptual coherence" in a manner similar to *prägnanz* (pp. 241f.).

16 The principle of Prägnanz is controversial. It is frequently criticized, especially from outside the Gestalt community, for being "vague … and subjectively defined" (Feldman, 2015, p. 1008). Others maintain that Wertheimer's definition of *Prägnanz* was clear and precise, but that it has been corrupted by later Gestalt theorists' uses of the term (Guberman, 2017). I will not problematize *Prägnanz* further, since the term is not used as a theoretical tool in this dissertation.

17 Reich, like others, erroneously attributes the quote to Schoenberg, which hardly alters the point.

Through both theoretical and empirical work, Wertheimer and the Berlin school of Gestalt psychology took Ehrenfels's concept further. A "Gestalt" is no longer understood as a summative quality, one-sidedly dependent on its parts, but as an organizational structure that often precedes and even influences our conscious grasping of the elements it contains (Wagemans, 2015, p. 5). This process might best be described through the concept of *emergence*, which offers an alternative to the premise of perception as a linear, additive combination of basic features. Rather, percepts may contain certain properties that arise from the juxtaposition of basic features in time and/or space, but that are not reducible to these: Pomerantz and Cragin (2015) maintain that "those properties—emergent features—behave as though they were elementary themselves, sometimes even being detected far more efficiently than the nominally more basic features from which they arise" (p. 88). Although emergent features *arise* from more fundamental percepts, they generally manifest a quality that is not inherently present in these fundamentals, but novel to them. This can be exemplified by the shapes of two triangles emerging from the juxtaposition of elements in the *Kanizsa Triangle* illusion (Figure 1).

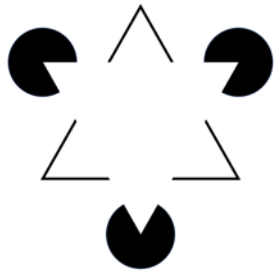


Figure 1: Kanizsa triangle (Wikimedia Commons)

A sort of auditory analogue to the Kanizsa Triangle is the beats that arise when two tones of similar but not identical frequency are played simultaneously, and which are used for tuning instruments (Pomerantz & Cragin, 2015, p. 97).[18] Another example is the quality that arises from sounding the pitches C, E♭, and G simultaneously. From the combination of these three pitches, a composite feature emerges that is not present in any of the individual tones: a quality we might call *minorness*. If we change the frequency of the E♭ pitch slightly so that it becomes E♮, the emergent feature of the configuration radically changes—into *majorness*. These examples all illustrate an aspect of emergence that sets them apart from perceptual objects without this quality: "Emergence also requires novelty, unpredictability, and surprise

18 It should be noted, however, that these beats are acoustical phenomena, i.e., amplitude modulations. In other words, the emergent feature of "beating" is a physical reality rather than a perceptual interpretation.

that make the whole *qualitatively* different from the sum of its parts" (Pomerantz & Cragin, 2015, p. 88). The examples also demonstrate that

> Gestalts, grouping, and EFs [emergent features] are inseparable concepts; when we say that two elements group, we mean that salient, novel features emerge from their juxtaposition in space or time. If a collection of elements contains no EFs, that collection is not a perceptual group. (Pomerantz & Cragin, 2015, p. 90)

I will adhere to this definition offered by Pomerantz and Cragin. In addition, I will sometimes use the term "Gestalt quality." Fundamentally, this is not different from the term "Gestalt": A Gestalt has a certain perceptual quality—a Gestalt quality—that makes it appear to the perceiver *as a* Gestalt. The merit of the term "Gestalt quality" is that it tries to capture this specific quality, rather than merely presupposing it.

Furthermore, I make the proposition that all conventional[19] chords (and many unconventional ones[20]) fulfill Pomerantz and Cragin's definition of being perceptual groups (i.e., Gestalts). Conventional chords possess salient emergent features that can be readily expressed through generic harmonic labels—e.g., "tonic," "Neapolitan," "A♭–7(♭5)," "power chord," "suspended chord," "chromatic mediant," "cluster," etc. These labels are all *holistic* descriptions, referring to an emergent perceptual quality—i.e., Gestalt qualities. For example, while "tonic" is a contextual term, it makes no sense to label any given pitch configuration as "tonic chord" unless it conveys a perceptual quality associated with this label.[21] In other words, labels like these refer to both harmonic concepts and to perceptual qualities. Some of these are simultaneous qualities, like power chords or sus4 chords. Others are sequential qualities, deriving their perceptual quality from juxtaposition in time, e.g., a Neapolitan chord. But in both cases, these perceptual qualities are *Gestalt* qualities, following Pomerantz and Cragin's definition. They arise from elements—individual tones or individual chords in a progression—that group together to form a novel feature which is more perceptually salient than the elements themselves.

To sum up, then, a Gestalt consists of parts that are perceived as a whole. In the perceptual grouping of these parts, new, salient features emerge that were not present in the parts. These

19 Such pitch configurations as would be the focus of harmonic ear training.

20 "Unconventional chords" is, of course, a large and staggeringly unspecified category of pitch configurations. I postulate it merely to refer to uncommon or unique pitch configurations that might not form salient emergent features (i.e., they are not heard "as" anything), and thus may have a weaker tendency to perceptual grouping. A more general version of this idea is expressed by Guberman (2015), who argues, with reference to Wertheimer, that "the short and simple description is a 'good' Gestalt, and the long and complex description is a 'bad' one" (p. 26).

21 E.g., stability, finality, centrality, arrival, or other qualities associated with "tonic-ness."

emergent features bring a specific perceptual quality to the Gestalt—what we refer to as its Gestalt quality.

2.4.2 Gestalt and subjective experience

At this point, I want to draw attention to an aspect of perception that has been less problematized by traditional Gestalt theory: the subjectivity of perceptual experience. While clearly acknowledging that the mind plays a role in perceptual grouping (Tan, 2015, p. 107), Gestalt theory has often focused its attention to the organizational principles *predicting* such grouping, i.e., the Gestalt laws. In other words, the emphasis has been on the nature of the percept, rather than on the perceiver.

We can exemplify this by taking another look at the Kanizsa triangle (Figure 1). The emergent feature of a white triangle covering another triangle and three black circles is largely predetermined by the arrangement of elements. It is conceivable that a hypothetical individual does not experience perceptual grouping when looking at the configuration, and instead describes the illustration as depicting three angle brackets and three black "pacmen." It is also conceivable that someone might see a *variation* of the white triangle, e.g., if the two overlapping triangles are grouped into a Star of David. It is unlikely, however, that someone would claim to see *radically other* emergent features, for instance a rectangle or an oval. The elements of the Kanizsa triangle simply do not support such groupings.

The perception of musical harmony, I propose, is much more divergent. This is in part because harmony is generally a richer and more complex perceptual object than visual figures, with both semantic, cultural, and emotional connotations. Harmonic *progressions*, furthermore, include a temporal dimension. While the basic grouping of individual pitches into harmonic Gestalts (i.e., chords) is relatively uniform and general, the degree to which an individual listener is attuned to musical harmony's various "meanings" is highly contingent on the listener's individual experience, knowledge, mental schemata, associations, cultural background, and so forth. The same chord may afford a range of reactions or interpretations in different listeners, while it is still justified to say they all perceive the chord itself as a Gestalt rather than as individual pitches.

One might object to my bringing subjective experience into the discussion on Gestalts. One may also object to the notion that Gestalt theory is concerned with "meaning" at all. However, I find such objections to be based on a misunderstanding. The interaction between stimulus and listener has been a presumption—implicitly, at the very least—in virtually all Gestalt-oriented studies of musical perception since Kurth's *Musikpsychologie* (1931). Gestalt

perception of music "is not merely a matter of musical structure, but is essentially dependent on structuring by the listener, since dealing with music is an active and constructive process" (Reybrouck, 1997, p. 61). This process is highly influenced by the listener's schemes and knowledge (Reybrouck, 1997, p. 60), which contribute to organizing the perceptual input in search of meaning (Reybrouck, 1997, p. 63). Furthermore, Gestalt theory and research have always been based on the experiences of individuals (Metz-Göckel, 2015, p. 22). This is not to say that "Gestalt is in the ear of the beholder"—I merely emphasize the dynamic relationship between percept and perceiver.

Precisely because these Gestalt concepts have shortcomings in explaining the *nature* of the relationship between perceiver and percept, however, I will eventually supplement them with other theoretical approaches to perception. For now, it should be sufficient to state that we do not exclude a subjective dimension in our definition of "Gestalt." On the contrary, we acknowledge that harmonic Gestalts are not static, objective, or unidirectional by nature. They are flexible and divergent perceptual objects that may sustain different Gestalt qualities in different listeners.

By allowing the term "Gestalt" to include the subjective, experiential quality it impresses on the perceiver, we simultaneously avoid having to introduce yet other terms for this dimension. For instance, a common philosophical term for subjective experience is *qualia* (singular: *quale*). While the term "qualia" have been used with different meanings, it is often used as a general descriptor of subjective experience (Schiavio & Van Der Schyff, 2016). In this understanding of the term, qualia refer, for example, to the quality of *what it is like* to taste an apple, to see the color red, to feel pain. However, adopting the term also means stepping into an old and heated debate, most of which is centered on "which mental states have qualia, whether qualia are intrinsic qualities of their bearers, and how qualia relate to the physical world both inside and outside the head" (Tye, 2021). While these questions are not avoided by simply choosing another label, the philosophical discourse surrounding qualia will be held outside this dissertation for practical reasons. Christopher Doll, in his monograph on stock harmonic patterns in anglophone rock and pop music (2017), also rejects the term qualia, because of its complexity. He uses instead "effect" and, synonymously, "quality" (Doll, 2017, p. 8). Other, more colloquial terms found in harmony literature include "color" (Gorow, 2006; Radley, 2008); "sonic color" (Mulholland & Hojnacki, 2013); "harmonic impression" (Harrison, 1995); "subjective effect" (Mulholland & Hojnacki, 2013); "how the sound makes us feel" (Coker et al., 1997); or "the chord's feel" (Karpinski, 2007).

I will use "Gestalt quality" (or simply "quality") to refer to the experiential dimension discussed here. In addition to the reasons already stated for using this term, it has the added benefit of

keeping the connection and relevance to Gestalt perception more explicit. However, "Gestalt quality," as used here, could more or less be substituted by "quale," were one to adhere to another philosophical tradition.

2.4.3 "Gestalt" vs. holistic harmony identification

There is still a critical problem that must be addressed: What is the relationship between a harmonic Gestalt as described here, and a harmonic Gestalt as it is portrayed by Karpinski? While this important distinction will be discussed later in the dissertation, it should be noted here that the two descriptions do not appear to be identical. *My* claim, leaning on a traditional Gestalt understanding and auditory scene analysis (cf. Bregman, 1990, p. 496), is that all conventional chords are Gestalts in the sense that they are generally perceived as singular acoustic objects. In fact, the inclination to hear chords as objects is something that usually has to be *unlearned* through aural training methods, such as identifying, singing, or arpeggiating their constituent pitches. In other words, hearing harmonic Gestalts in the more traditional understanding of the term is not contingent on any special skill or expertise. A basic perceptual preference for grouping pitches into objects may simply be a natural tendency in the human auditory system.[22] Thus, using a traditional Gestalt understanding, I maintain that pitch simultaneities are generally perceived as Gestalts, just like melodies are.[23] The same goes for harmony as a successive phenomenon, i.e., in a chord progression: "Harmony, just as melody and rhythm, is comprised of patterns and generally is perceived holistically. Individuals respond to harmony as a totality rather than as individual tones or chords" (Radocy & Boyle, 2012, p. 239).

Karpinski's *Gestalt listening*, however, goes much further: It presupposes recognition and identification, an understanding in some sort of musical way of *what* one hears.[24] In his version, hearing harmony as Gestalts is an acquired expertise. These Gestalts present themselves perceptually to some, but not to others. Chenette et al. (2021) use this paradox to question whether what Karpinski describes can be considered true Gestalts, since Gestalts are supposedly

22 Note that the perceptual pitch grouping described here should not be confused with tonal fusion (or—with a more recent term—*harmonic* fusion), which refers two or more to individual pitches being perceived *as one pitch*. In a harmonic Gestalt, the constituent pitches do not have any particular inclination towards being perceived as one pitch. They simply form an integrated whole—a Gestalt, a chord—that possesses a new quality that is not inherent in the parts.

23 Expanding on Ehrenfels's example of a melodic Gestalt, Wertheimer points out that upon hearing a melody (17 tones) with its accompaniment (32 tones), we do not hear "49," and certainly not "20 plus 29," but simply "melody" plus "accompaniment"—even when performed in a way that (in more modern terms) impedes auditory stream segregation (Wertheimer, 1923/1938, p. 71).

24 While Karpinski mentions the ability to express this understanding through theoretical labels (e.g., "subdominant chord"), I think we should not exclude non-theoretical forms of identification. For instance, a country guitar player may be intimately familiar with the Gestalt quality of a IV chord, without knowing the term "subdominant."

perceived immediately, while Karpinski's skill is only acquired after extensive exposure. I do not share this concern. Gestalts may be lower- or higher-level, and there are degrees of expertise involved in most all listening. Consider the audio engineer making minute EQ adjustments, or the orchestra conductor quietening the cellos in order to balance the sound of the string section. These are reactions to imperfections in the overall sound (i.e., Gestalt), but both the detection and diagnosis of the imperfection require expertise.

The factor of expertise can also be illustrated by a different comparison: that of listening to a spoken language that one does not understand, versus listening to a language one understands. Both are perceived holistically, that is, as "speech" rather than as detached, unrelated sound bits. But whereas the unfamiliar language is perceived as an auditory stream of phonetic and inflectional variation (or perhaps, depending on the listener's experience, with a somewhat more specific Gestalt quality, e.g., "some Slavic language"), the familiar language is perceived as clear words and sentences. The speech is no longer perceived primarily as *sound* at all, but as *meaning*.[25] Learning a new language will alter how this language is perceived. While the language analogy should not be stretched too far, Karpinski's concept has some clear parallels: His *Gestalt listening* is about perceiving harmony as musical meaning rather than merely as musical sound. This indicates that it involves an element of perceptual learning, even expertise. We might also propose that Gestalt listening is characterized by top-down, concept-driven processes, whereas the basic perceptual grouping of individual pitches into an auditory object (i.e., "chord") is primarily a bottom-up process. However, the difference in listener capacities that separates these two perceptual modes must eventually be examined closer.

Let us conclude for now that the term *Gestalt* will henceforth be used in the more traditional sense outlined above. Karpinski's use of the term seems to be less precise and has a somewhat enigmatic component to it (which it is the ambition of this dissertation to elucidate). To avoid confusion, I will henceforward refer to Karpinski's concept as *holistic harmony identification* rather than *Gestalt listening*. Karpinski himself uses the term *holistic identification* elsewhere to describe the same phenomenon (Karpinski, 2007, p. 157), so this choice seems justified. Apart from avoiding confusion over the specific meaning of the term *Gestalt*, *holistic harmony identification* has several advantages. Without excluding traditional Gestalt theory approaches or diminishing their relevance, it will let us start off this enquiry from a less dogmatic and potentially alienating theoretical position. Furthermore, *holistic harmony identification* does not risk creating the impression that it is some sort of Gestalt theory-specific "technique." Most importantly, however, it accentuates the concept's main point: that it involves *identification*

25 Incidentally, Clarke (2005, p. 34) points out that the acoustical features of language are more easily noticed by a listener who doesn't understand it, precisely because the native speaker is much more concerned with the *meaning* expressed than with its distinctive features qua *sound*.

and not merely "listening." This will help us shift the focus from the general auditory system to the individual, phenomenal experience.

2.4.4 Conclusion

In spite of its musical origins (cf. Ehrenfels), Gestalt theory has been predominantly focused on visual experience. Translating insights from this research to the domain of music, then, might be considered speculative (Reybrouck, 1997, p. 64). Gestalt theory's shortcomings as an explanatory model must also be acknowledged. A pragmatic and common approach to this theoretical uncertainty is to use Gestalt theory as a heuristics for studying perceptual organization, without adhering to its more dogmatic stances or relying on its explanatory power. Following this example, I will use the Gestalt concepts of "Gestalt" and "emergence" to conceptualize and highlight aspects of holistic harmony perception—whatever it may be—that might otherwise remain unarticulated.

3 A perceptual framework for exploring Karpinski's Gestalt listening: A holistic, ecological approach

3.1 Introduction

What kind of perceptual theory can best explain holistic harmony identification? "Gestalt theory" might seem like an obvious answer to this question. Yet as we have seen, the Gestalt movement has been extensively criticized for focusing more on descriptive principles than on explaining perceptual processing (see, e.g., Bruce et al., 1996, p. 110).

There are at least two aspects to this criticism. One might of course take it at face value, acknowledging that while Gestalt theory has convincingly described the *premises* for perceptual grouping—cf. the so-called "Gestalt laws" first formulated by Wertheimer (1923/1938)—it does a poorer job of explaining how this perceptual information is actually processed by the brain. On the other hand, one might argue that the criticism is partially flawed, because it is raised from within a theoretical paradigm that presupposes "information-processing" in the first place. While an information-processing understanding was the paradigmatic approach in perceptual psychology for at least the last few decades of the twentieth century (Clarke, 2005, p. 11), alternatives to this view have since gained traction. I will make the argument that an information-processing approach is not the best suited perceptual framework for exploring holistic harmony identification. I will then present an opposing framework which seems to highlight some of its central facets.

3.2 Weaknesses in an information-processing approach

What Clarke (2005) refers to as an "information-processing" approach to perception—and which one could also describe as a "conventional" cognitive approach—is a very simplified presentation of a range of theoretical positions, not necessarily in complete agreement with each other. What they have in common, however, is that they regard the environment as

> a maelstrom of sensory stimulation which perceiving organisms organize and interpret with the processes and structures of their perceptual and cognitive apparatus.

> In other words, structure is not in the environment: it is imposed on an unordered or highly complex world by perceivers. (Clarke, 2005, p. 12)

A more recent portrayal by Hatfield (2014) offers more nuance:

> The main commonalities involve the notion that cognition is information processing that explains intelligent behavior. The differences concern whether early perceptual processes are cognitive, whether representations are needed to explain cognition, what makes something a representation, and whether cognitive processes are limited to the nervous system and brain or include other bodily structures or the environment itself. (Hatfield, 2014, p. 361)

Clearly, this is not one thing. Nor is it a static tradition—which is exemplified by the substantial attention given to the role of multimodality and/or embodiment in perception research in recent years (e.g., Godøy, 2017, 2018a, 2018b; Gordon et al., 2018; Iyer, 2002; Leman & Maes, 2014; Stiles & Shimojo, 2015; see also the anthologies Shapiro, 2014, and Wöllner, 2017). Thus, it is not my intention to give a distorted or caricatured presentation of conventional cognitive approaches. Still, leaning on Clarke's (2005) simplified version, I will point to some weaknesses in an information-processing account of perception—and in particular, how these weaknesses pertain to the subject of holistic harmony identification.

Clarke offers a schematic representation of the basic outlines of an information-processing approach to music perception (see Figure 2), which illustrates the cognitive process proceeding from simple, physical stimuli, through gradually more complex cognitive structures.

While Clarke's illustration may give the impression that information-processing accounts of perception are unidirectional, involving only bottom-up processing, he acknowledges that this is not the case. Generally, such models accept a reciprocal interaction between top-down and bottom-up processing,[26] and the former generally account for individual differences in perceptual interpretation. However, the perceptual process is still seen as "working primarily from the bottom up … with more complex levels constructed from the outputs of lower-level, more primitive processes." (Clarke, 2005, p. 15) Arguing against this premise, Clarke makes the claim that if one hears a snippet of music on the radio, one is likely to be aware of genre (a "high-level" feature) much more immediately than the "low-level" features (such as pitch intervals, key, meter, or instrumentation) that supposedly would have to be processed first (Clarke, 2005, p. 16). This is a point that, since Clarke's book, has been demonstrated empirically

26 This is also supported by recent research on music and the brain: A meta-analysis by Pando-Naude et al. (2021) of 130 neuroimaging studies found that music perception must involve both top-down and bottom-up processes.

by Gjerdingen and Perrott (2008). In a study on the recognition of music genres, they found that participants were exceedingly able to correctly match short snippets of music with one of ten genre labels. Even for the shortest excerpts, at a mere 250 milliseconds, participants scored at an above-chance level. This illustrates that the perception of musical meaning is not inherently additive, and that information-processing approaches fail to convincingly account for Gestalt-like perception.

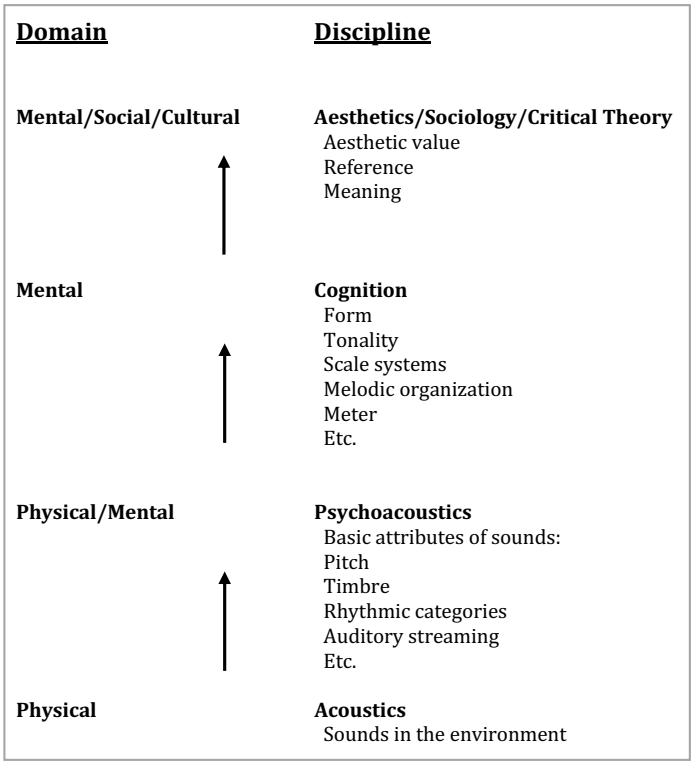


Figure 2: Clarke's (2005) schematic representation of an information-processing approach

A Gestalt is, purportedly, grasped instantaneously and without conscious perceptual analysis. The perceptual "summing" of elements described by computational theories is precisely what Gestalt psychology opposes. Regardless of how we eventually define holistic harmony identification, we take for granted that it is characterized by the same immediateness and spontaneity as is described by Gestalt perception, and by the same lack of perceptual analysis. Thus, we see the unsuitability of exploring holistic harmony identification with an information-processing, hierarchical approach to perception. It has also been argued that traditional cognitive theories

are generally unable to account for important nonlinguistic aspects of music behavior, such as its emotional and associative qualities (Iyer, 2002).

On a side note, a staple of the traditional approach to harmonic listening methodology mimics an information-processing cognitive model: Students are encouraged to analyze harmonic passages by synthesizing lower-level information, e.g., arpeggiating chords, listening for the presence of added tones or the tritone interval, or listening for the bass line in relation to a key center. While it is far from my intention to suggest these practices have no merit,[27] it is not obvious that they encourage a holistic listening approach.[28]

Rather than adhering to an information-processing paradigm, I will sidestep its self-contradictory implications for holistic harmony identification altogether by adopting a radically different approach: the ecological approach[29] established by James Gibson (e.g., 1966)[30] and later accommodated to a musical context by (among others) Clarke (2005).

3.3 Ecological theory

3.3.1 Gibson's affordances

Like the Gestalt psychology movement of which he had once been a part (Norman, 2013, p. 12), Gibson opposed the view that perception involves mentally arranging and interpreting a chaotic and unorganized external world. Instead, he proposed that perceptual information is already structured in the environment:

> Instead of supposing that the brain constructs or computes the objective information from a kaleidoscopic inflow of sensations, we may suppose that the orienting of the organs of perception is governed by the brain so that the whole system of input and output resonates to the external information. (Gibson, 1966, p. 5)

27 On the contrary, such strategies can be important for perceptual learning, i.e., learning to detect features in the stimulus information that were previously undetected (cf. Clarke, 2005, p. 24).

28 Interestingly, Karpinski declares that holistic harmony identification is "difficult to develop directly," but discusses only non-holistic listening strategies in his chapter on harmonic dictation (Karpinski, 2000, p. 119).

29 Sometimes referred to as Gibsonian psychology.

30 While Gibson is widely considered a founding figure of the ecological approach, Reybrouck (2021) traces its roots back to Haeckel (1866).

"Resonating" to the external information is not a passive process, notes Clarke: Perception is fundamentally *exploratory*, and "tuning in" to the perceptual information in our surroundings is closely linked to an active engagement with this environment (Clarke, 2005, p. 19). This way, perception is also closely related to *meaning*: "When people perceive what is happening around them, they are trying to understand, and adapt to, what is going on. In this sense they are engaged with the meanings of the events in their environment" (Clarke, 2005, pp. 6–7).

Chemero (2009) sums up Gibson's ecological approach as having three major tenets:

> First, perception is direct, which is to say that it does not involve computation or mental representations. That is, Gibson thought that perception was not a matter of internally adding information to sensations. Second, perception is primarily for the guidance of action, and not for action-neutral information gathering. We perceive the environment in order to do things. The third tenet follows from the first two. Because perception does not involve mental addition of information to stimuli, yet is able to guide behavior adaptively, all the information necessary for guiding adaptive behavior must be available in the environment to be perceived. Thus the third tenet of Gibson's ecological approach is that perception is of *affordances*, that is, directly perceivable, environmental opportunities for behavior. (Chemero, 2009, p. 23)

Affordances,[31] in ecological theory, are the possible interactions between environment and perceiver. They represent the relationship between the needs and capacities of an organism and the opportunities that are available in the environment (Clarke, 2017, p. 528). Central to the idea of affordance is precisely the reciprocity between perceiver and percept. Affordances are not static features, but products of this interrelationship. What matters is "not merely the world in its objective qualities, but the world as perceived by organisms" (Reybrouck, 2012). For instance, ecological theory would agree with Gestalt theory that a brick wall is not grasped by perceptually "adding together" the individual bricks from which it is composed. Before we register such details as the size, shape and arrangement of the bricks, the color of the mortar, etc., we simply perceive the concept "wall." Where the two traditions differ, is in ecological theory's emphasis on what "wall" *means* to the perceiver, and what it affords. As a physical barrier, it might afford stopping, walking around it, or keeping out. But to a child, it might afford climbing, balancing, or kicking a ball against it. To a deer being chased by a predator, it might afford jumping. To a dog, it might afford sniffing and leaving a territorial mark, or taking a nap in its shade. In this perspective perception, meaning, and action are

31 "The verb *to afford* is found in the dictionary, but the noun *affordance* is not. I have made it up. I mean by it something that refers to both the environment and the animal in a way that no existing term does. It implies the complementarity of the animal and the environment." (Gibson, 1979, p. 122)

closely connected: Objects in the environment become meaning-carriers when they enter into a relationship with a subject (Reybrouck, 2021, p. 73).

This further reveals two important aspects of ecological theory: First, it directs attention to the fact that an organism's activity in relation to the surroundings *generates* stimulus information (Hatfield, 2014, p. 365). Second, it illustrates that perception is not regarded as a mentalistic, abstract activity, but as fundamentally embodied and multimodal (Norman, 2013, p. 12). The emphasis on the concrete rather than the abstract is one of the main advantages of an ecological approach over a conventional cognitive approach (Parncutt, 2009, p. 127).

Lastly, a word on perceptual learning in a Gibsonian perspective. Conventional cognitive psychology tends to explain the enhancement of an individual's perceptual capacities in terms of increased "coding power" and an accumulation of knowledge (Clarke, 2005, p. 22). Because ecological theory maintains that the environment is already structured—rather than a chaos that must be "sorted out" mentally—it views perceptual learning as an increasing ability to distinguish and differentiate between these structures (Gibson, 1969; Gibson & Gibson, 1955).

3.3.2 Criticism of "cultural" affordances

There has been disagreement on how Gibson's concept of affordances translates to a cultural context, e.g., whether they should apply to complex, human-made artifacts or only to natural properties (Windsor & de Bézenac, 2012, p. 104). Windsor and de Bézenac (2012) summarize the discussion thus:

> At one extreme are those who view cultural perception as requiring a switch to more traditional thinking on inferential cognition (e.g., Chow, 1989; Reed, 1991); at the other are those who argue that direct perception for affordances can (and should) account for perception and action in complex cultural contexts (e.g., Noble, 1991; Sanders, 1997; Heft, 2001; Windsor, 2004). (p. 105)

Gibson himself was ambiguous on this point, although he specified that "the richest and most elaborate affordances of the environment are provided by other animals and, for us, other people" (Gibson, 1979, p. 135, quoted from Windsor & de Bézenac, 2012). Windsor (2004) suggests that Gibson indirectly touches upon cultural interpretation also in another context: In cases of "inadequate information"—e.g., information that is "conflicting, masked, equivocal, cut short, reduced, or even sometimes false" (Gibson, 1966, p. 303)—the perceptual system "hunts" for additional information that will reinforce one interpretation or the other. This, argues Windsor, includes social and cultural information (2004, pp. 194–195). Windsor's

view suggests that "aesthetic objects afford interpretation because they deny the possibility of exploration with the other perceptual senses one would normally use to make sense of information" (Dibben, 2001, p. 184).

Needless to say, proponents of using an ecological approach to examine music perception tend to lean towards the cultural affordances camp. While the "cultural" approach can be said to depart from an established understanding of ecological principles, it can arguably also be seen as an adaptation and elaboration of these principles to the environments of modern humans.

3.3.3 An ecological approach to music perception

Gibson and his followers are mostly concerned with visual detection of affordances (Krueger, 2014, p. 2). Later ecological psychologists have concentrated on auditory events (see Windsor & de Bézenac, 2012, p. 105 for a review of this research). Most relevant to the context of music is the work of Gaver (1993a, 1993b), who distinguishes between "everyday listening" and "musical listening" (Gaver, 1993b). Following Gaver, there has been a significant interest over the last decades in applying an ecological approach or the concept of affordances to music perception (e.g., Christensen, 1996; Clarke, 2005, 2012; Cook, 1998; Dibben, 2001; Godøy, 2010; Kozak, 2020; Krueger, 2011, 2014; Menin & Schiavio, 2012; Nussbaum, 2007; Reybrouck, 2005, 2012, 2015; Ruud, 2020; Windsor & de Bézenac, 2012).

Music is obviously a more complex perceptual object than, say, a brick wall. It is also usually embedded in some kind of social context. For these reasons, Gibson's concept of affordances—which was primarily *action*-oriented—has generally come to include more nuances. While an ecological take on music perception might very well focus on physical response, e.g., various aspects of entrainment, a more introspective approach is also possible: Krueger (2014) proposes that "when we integrate with music via musical affordances … we are able to temporarily access musically scaffolded forms of experience and expression that we cannot access outside of this music-listener system" (p. 8). Some go even further: Musical affordances, according to Clarke (2017), are "what people perceive they can do with, or make of, the music that they encounter, whether as performers or listeners" (p. 528). This includes much more than just overt action. In fact, states Clarke, there is no such thing as "passive listening" (2005, p. 205). Thus, Clarke suggests that *interpretation* is also a form of action, and, followingly, that musical affordances in this sense are closely connected to the perception of musical *meaning* (p. 204). This view takes into account the social nature of affordances for humans: "For human societies, with all the complexity of their cultural elaboration, affordances can become hugely ramified and elaborated, even if the more basic affordances are never left behind" (Clarke, 2017, p. 528).

Clarke has been criticized by Menin and Schiavio for using the concept of musical affordances too broadly and without conceptual consistency, to the point where "the meaning of *to afford* … does not exceed the colloquial concept of *to evoke*, or *to elicit*, showing that musical experience is, in a vague sense, evocative" (Menin & Schiavio, 2012, p. 206). Those who, like Clarke, take a more inclusive stance, maintain that while musical affordances may be idiosyncratic, they are not random or completely subjective, but constrained by material properties (e.g., Dibben, 2001; Windsor & de Bézenac, 2012). Dibben (2001), adapting an argument by DeNora (2000), suggests that "sounds specify meanings and values for particular listeners, some or one of which can be *mobilised* [emphasis added] at particular moments" (Dibben, 2001, p. 184). This understanding, in my view, acknowledges the idiosyncrasy of musical affordances without undermining the fundamental ecological principle of object–subject reciprocity.

In this dissertation, I will follow the more inclusive and culturally oriented understanding of affordances. While this choice may be criticized for being speculative, it is not primarily intended to address the general debate on the nature of affordances. Rather, I choose it as a practical and heuristic approach, making use of ecological principles to "frame [a discussion] of how listeners make sense of music" (Clarke, 2017, p. 528)—specifically, how we make sense of harmony.

3.3.4 Discussion on the role of musical affordances in this dissertation

It is not my intention to dismiss decades of music cognition research, or to portray a conventional cognitive approach to music perception as erroneous or invalid. However, I believe that the conventional approach has limitations in accounting for the spontaneous, holistic harmony identification that is the topic of this dissertation.

This partly concerns the act of holistic perception itself. Proponents of a more conventional approach may argue that what I here refer to as "holistic perception" is really just information-processing that is so effective as to appear as instantaneous. That is an ontological discussion which is outside the scope and aim of this dissertation. We are primarily concerned with Karpinski's term "Gestalt listening" understood as a *symptom* of a certain type of harmonic identification, which appears to the listener as instantaneous and non-analytic. Therefore, we should entertain the possibility that this is an umbrella term for perceptual experiences that are not ontologically identical, but which nevertheless have a similar practical outcome: a "rapid and fluid" identification of harmony (Karpinski, 2000, p. 119).

More importantly, I find that conventional music cognition has important limitations when it comes to the epistemological and phenomenological aspects of holistic harmony identification.

Far more pedagogically relevant than what holistic harmony identification objectively "is"—which might not even be answerable—are questions like how it manifests itself to the listener, how it is impacted by listener intentionality, and what the nature is of the interaction between percept and perceiver in such identification.

Thus, Gibson's idea of *affordances* fills an explanatory weakness in both Gestalt theory and conventional music cognition. It conceptualizes the dynamic relationship between perceiver and environment. The elaboration of Gibson's concept to a music perception context offers a framework for examining which interpretations musical structures may or may not afford. In a discussion of holistic harmony identification, it may help us verbalize the acts of meaning detection involved in such perception, and how harmonic Gestalts can afford different meanings to different listeners in different contexts.

An ecological approach to music perception further avoids two positions that both seem incapable of providing a convincing account of the nature of holistic harmony identification: On the one extreme, the position of complete musical autonomy, whereby all relevant perceptual information or "meaning" is seen as inherent in the musical structures themselves. Musicology has often been criticized for adopting—or taking for granted—such a view of musical meaning (e.g., DeNora, 2000; Small, 1998).[32] And, on the other extreme, a total subjectivism, whereby all meaning is viewed as a construct of the empirical listener. Naturally, these two positions are not the only possible positions, but rather the poles of a continuum. An ecological approach, nonetheless, in a way seeks to bypass the whole continuum by proposing a more dynamic perceptual relationship between musical structures themselves, their acoustical properties, the listening environment, social and cultural contexts, and the empirical listener's personal experience, predispositions, and capacities.

It is this dynamic relationship between stimulus, listener and environment that prompts me to complement the Gestalt concepts discussed in the previous chapter with the concept of affordances. However, the fusion of Gestalt principles with an ecological approach to music perception is not so obvious as to be accepted without scrutiny. Therefore, we should examine whether these two approaches to perception are logically compatible.

32 Clarke, while critical to the tradition of musical autonomy, suggests a practical approach to the listening practices with which it is associated: "Music can be understood as constituting a 'virtual world' into which a listened is drawn. An ecological approach can then help to understand the ways in which the events of this virtual world (tonal events, metrical events, textural events, motivic events) are specified in sound" (Clarke, 2005, p. 154).

3.4 Compatibility of ecological and Gestalt approaches

Gibson himself had a background in Gestalt psychology (Norman, 2013, p. 12). Evidently, he found this tradition lacking in some regard, and he went on to develop his own approach to perception. Still, the two theoretical strands share a number of views. Let us examine these closer.

First and foremost, both oppose additive or information-processing accounts of perception. They both make the claim that structure (at least partly) exists in the environment, and that the observer perceives these structures directly. Ecological theory advocates an embodied and multimodal view on perception. The embodied perspective has—although perhaps less explicitly—always been at the core of Gestalt theory, too (Guberman, 2016). So has the interest in intersensory Gestalts, although there is as yet little scientific evidence of their existence (Spence, 2015; Spence & Di Stefano, 2022).[33] Furthermore, both theories are also concerned with meaning as *part* of the perceptual process.

Where they depart somewhat from each other, is in their understanding of the relationship between object and perceiver. While they both acknowledge an interrelationship, Gestalt theory is less elaborate on its nature. Its focus is primarily on object properties, and less on the perceiver capacities that precipitate their pick-up. Gestalt theory, in my understanding of it, also has limitations in accounting for individual differences and the role of learning. In ecological theory, the interaction between object and perceiver is more explicitly conceptualized, through the concept of affordances. Ecological theory also emphasizes the importance of perceptual learning and social context.

Likewise, Gestalt theory's concentration on the prerequisites for holistic perception offers valuable perspectives on stimulus information that are less developed in ecological theory. The concept of emergent features is an example of this: It offers a framework for understanding how combinations of pitches may afford such remarkable perceptual interpretations as "longing," "anguish," "romance," "jazz," "cluster," etc.

Rather than expressing an incompatibility between the two theories, these differences in focus form my primary argument for combining them: Each of the two theories elaborates an aspect of perception that is less developed in the other.

33 Although a multimodal understanding of perception is not limited to these two theoretical traditions—it is arguably also becoming part of a mainstream cognitive approach: "Multimodality has in recent years received a lot of attention in the cognitive sciences, and 'classical' notions of the separation of the senses have been challenged. There is now mounting evidence that the sense modalities work together and complement one another" (Godøy, 2017, p. 8).

In conclusion, I find no inherent contradictions in a Gestalt-oriented ecological framework. On the contrary, there is substantial overlap between the two. Yet, there is a gain in combining them: As already discussed, I consider ecology to offer a better explanation of the interaction between stimulus, listener, and environment (i.e., affordances) than both conventional cognitive approaches and Gestalt theory. Thus, it fills some of the explanatory gaps in Gestalt theory regarding how the perceiver's individual capacities resonate to the structural properties of the percept, embedded in a social context.

3.5 Other theoretical approaches

The Gestalt-oriented ecological approach presented here will form the overarching framework for the theoretical examination of holistic harmony identification. Additional theoretical approaches will be introduced more contextually to supplement, comment, or contrast this framework.

4 What does it mean to identify harmony holistically?

4.1 Introduction

In the theory chapter I coined a new term for Karpinski's concept of "Gestalt listening": *holistic harmony identification*. This was done partly to avoid confusion with the general perception of chords as Gestalts, and partly to emphasize important facets of the concept (such as *identification*). However, introducing a new term means introducing new words with different connotations and meanings. Therefore, as part of our theoretical exploration of holistic harmony identification, we should investigate the notions included in the three words *holistic*, *harmony*, and *identification* in this context, and following on from that, try to explicate the precise meaning of their combination.

4.2 What is the meaning of *identification*?

In the paragraph describing Gestalt listening, Karpinski uses the word *recognize*[34] eight times. This underlines the aspect of familiarity, of pinpointing a perceptual quality that is already known to the listener. In cognitive psychology, *recognition* is understood as "the automatic activation of some particular contents of long-term memory that have some relation or association with current perception" (Snyder, 2000, p. 10). However, recognition in this sense is not sufficient for the kind of listening Karpinski describes, which also involves *knowing what one hears*. Elsewhere, Karpinski uses the word *identification* to encompass this meaning,[35] though without distinguishing it clearly from simple recognition. A closer definition may again be found in Snyder (2000): "*Identification* occurs when we not only recognize something, but are able to connect it with memories of its name and associated with its concept" (p. 10).

Still, this definition warrants some further nuancing when used in the specific context of aural training. The first clarification concerns the "what." Exactly *what* is identified in aural training is important. In an everyday context, we might say that recognition corresponds to the feeling "I've heard this song before," while identification is "this is Keith Jarrett's *My Song*." In

34 Or the variants *recognizing*, *recognizable*, or *recognition*.

35 E.g., later in *Aural skills acquisition* and in Karpinski, 2007 (p. 157).

a harmonic dictation, however, identifying the music as Keith Jarrett's *My Song* is beside the point at best. The relevant information to identify is the song's harmonic properties.

The second clarification concerns the "how," and this is somewhat more complex. *How* is something identified? The first part of Snyder's definition states that to identify something is to "connect it with memories of its name." This indicates a declarative form of knowledge: To aurally identify a given chord is to be able to state its music-theoretical label, either in the form of chord quality (e.g., "major triad in first inversion"), chord symbol (e.g., "E♭/G"), or harmonic function (e.g., "Neapolitan subdominant"). Academia has often been criticized for exclusively valuing such declarative knowledge, ignoring procedural forms of knowledge (e.g., Johannessen, 2013). The second part of Snyder's definition, however, adds "memories … associated with its concept." This is somewhat vague, but seems to consider forms of knowledge beyond that which can be stated as a fact. Many theoretical dichotomies describe such forms of knowledge, e.g., *epistêmê* vs. *technê* (e.g., Aristotle, 2020), knowing-that vs. knowing-how (Ryle, 1949), explicit knowledge vs. tacit knowledge (Polanyi, 1967), or analytical vs. intuitive knowledge (Swanwick, 1994).[36] DeBellis (1995, 2005) offers a more aural-training-relevant dichotomy: conceptual vs. non-conceptual listening. Perceptual concepts, he points out, are a function of what a person can recognize and discriminate (DeBellis, 2005, p. 56). A listener without musical training may, for example, lack a perceptual concept for scale degree $\hat{5}$ altogether. Such a listener will not have the ability to discriminate perceptually between $\hat{5}$ and non-$\hat{5}$, or to perceive $\hat{5}$s as being similar to each other. This, DeBellis argues, goes to show that aural training is "first and foremost the acquisition of a certain repertoire of perceptual concepts for musical events and properties, where one initially has no perceptual concepts of them" (DeBellis, 2005, pp. 55–56). If this were not the case, there would hardly be a need for aural training. The most interesting question posed by DeBellis, however—relevant to the present discussion—is what the relation is "between a trained listener's perceptual concepts and the music-theoretical concepts in terms of which he reports his hearing?" (DeBellis, 2005, p. 57) It is not difficult to imagine a listener who has a perceptual concept for scale degree $\hat{5}$, but who does *not* have a music-theoretical concept for expressing this. For example, young players who have learned music through playing by ear may fit such description. Should we state that this hypothetical listener has the ability to *identify* scale degree $\hat{5}$, according to Snyder's definition? Or simply to *recognize* it? For clues, we might look to how knowledge is generally viewed in aural training.

For pedagogical reasons, an aural training setting normally requires the student to *demonstrate* that she has indeed identified the properties of the sound that are relevant to the situation. This means that she must be able to express her identification through some sort of response.

36 Although, as I will eventually return to, Swanwick's terms are not really dichotomous, but hierarchical.

If we look to general aural training in the Western world, such a response may traditionally take many forms, according to curriculum, type of activity, and pedagogical convention: written notation, solmization, playing, clapping, takadimi, solfège hand signs, or verbal communication.[37] This list—which is not exhaustive—demonstrates that both knowing-that and knowing-how are considered relevant forms of knowledge: Being able to play back something correctly on your instrument is also a valid response.[38] In this view, it seems fair to say that our hypothetical listener does in fact have the ability to identify scale degree $\hat{5}$. As long as he is able to convincingly demonstrate this identification, e.g., through an instrument-based response, the term *identification* is not reasonably contingent on whether he also possesses the ability to express his understanding through a music-theoretical concept. For an even clearer example, we might consider a guitar player in a pop/rock band who is able to perfectly reproduce the chord progression of a song after listening to it once. It does not seem reasonable to say that this guitarist has not identified the chord progression, simply because she lacks the ability to offer a symbolic representation of it, e.g., through chord symbols, Roman numerals, Nashville numbers,[39] or harmonic functions. This pragmatic view is supported by Pratt, who states that "the greater need is to hear what a musical phenomenon *is* rather than to know what it is *called*" (Pratt et al., 1998, p. 5).

The argument I'm making here partly corresponds to both Karpinski's (1990, 2000, 2021) model for music perception during melodic dictation and Sloboda's (2005) four-stage theory of musical awareness. The first three stages[40] of both models deal with how relevant aspects of sound are attended to, processed, and finally "understood" (i.e., identified). The fourth stage of both models describes the production of a response. Since Karpinski's model is explicitly

37 "Mimicking" is not included in this list, e.g., singing back a melody one has just heard, or imitating a rhythm. Simply repeating what has been heard is a test of musical memory rather than of musical understanding, i.e., identification (Karpinski, 2020). Singing back a melody using solmization, however—or playing it back on an instrument—involves a transformation and reproduction of the perceived sound and should not be considered mimicking.

38 Aural training is not uniform across Western music institutions (or even necessarily within a single institution). Therefore, there may be opposing views to the argument I am making here—for example in institutions which favor a more traditional emphasis on sight-singing, dictation, and *knowing-that*. However, I believe it is a fair assessment to say that aural training in the Western world over the last decades has generally evolved towards an integrated approach to musicianship, and supplying students with metacognitive strategies for real-life situations. This view is voiced by, e.g., Cleland and Dobrea-Grindahl (2013, pp. 83–84) and Stillie and Moir (2021), but is also reflected, e.g., in the course description for *Aural training II* at my own institution. The course description explicitly mentions the use of one's own instrument in class, and learning objectives include "be[ing] able to transfer skills acquired on [*sic*] the aural training course to [one's] music practice," and "master[ing] different aural strategies in order to solve musical challenges" (Norwegian Academy of Music, 2021, para. 2). While it is conceivable that this shift is more pronounced in Scandinavia than elsewhere, Francis (2021) reports a similar intention in aural training at the Royal College of Music, London.

39 Nashville Number System: A shorthand and simplified notation system for harmonic progressions, resembling the Roman numeral system but using Arabic numbers. Compared to chord symbols, Nashville numbers have the practical benefit of not specifying a key, making it very flexible in studio sessions. See, e.g., Williams (2019).

40 Or "phases," in Karpinski's model.

concerned with melodic dictation, it specifies a response in the form of notation (Karpinski, 2021, p. 43). Sloboda is less explicit, and simply holds that the listener must be able to translate the received information into a response (Sloboda, 2005, p. 176). The response must, nonetheless, be relevant: A response in the form of clapping will usually not be relevant in a harmonic dictation, just like chord symbols are not relevant for a rhythmic dictation. Aural training activities will usually specify the relevant form(s) of response.

The important point is that the response is viewed by both Karpinski and Sloboda as an integral part of the act of perception—as is also maintained by ecological theory. It does not merely serve as an external goal (i.e., offering the teacher a chance to detect faults in the student's understanding, or supplying a measure of achievement which can be assessed and graded), but represents a transformation of the perceived sound into a corresponding expression. This transformation presupposes an *understanding* of the relevant musical properties—e.g., metric grouping, relative rhythmic proportions, or tonal function (Karpinski, 2000, p. 78).

In light of the foregoing discussion, let us revisit Snyder's definition of "identification" to see whether it might be refined to better match the context of aural training. The definition originally read: "*Identification* occurs when we not only recognize something, but are able to connect it with memories of its name and associated with its concept." We are now able to add some clarifications regarding *what* is identified, and *how* it is identified.

The *what* are the relevant musical properties of the sound, as specified and confined by the pedagogical activity at hand. Other properties may be indirectly relevant to the task—e.g., recognizing a piece or genre may elicit a concept-driven utilization of previous knowledge which strengthens the identification process—but they are not pertinent to identification as such.

The *how* has two important aspects: the nature of our knowledge, and how this knowledge is expressed. Snyder writes "memories of its name *and* [emphasis added] associated with its concept." We have interpreted these two categories as declarative and non-declarative knowledge respectively. While aural training arguably aims at making non-declarative forms of knowledge declarative, the two are generally treated as equally important categories: *knowing-how* is just as essential as *knowing-that*. Therefore, the word "and" in Snyder's definition should be replaced by "or." Furthermore, the *how* of identification in aural training demands that the knowledge be activated through a relevant external response. The response is both a manifestation and a translation of the knowledge. It is the action that is afforded by the perceptual environment, i.e., the aural training setting.

Building on these clarifications of Snyder's definition, I put forward the following definition of "identification" in aural training:

To identify means to detect, recognize, and differentiate relevant properties of a musical percept, and to be able to express these properties through a corresponding, external response.

4.3 What is the meaning of *holistic*?

As I have already argued, with reference to auditory scene analysis, the perception of harmony is generally holistic: We tend to perceive chords as Gestalts, not as simultaneous but unrelated pitches. What we refer to as a "chord" is the emergent feature of perceptual pitch grouping. The ability to separate and identify all the constituent tones of a chord is something one generally must learn through aural training. We might refer to this aspect of harmonic Gestalts as the holistic *perception* of harmony. Holistic harmony *identification*, as was discussed in Chapter 2.4.3, differs from this. With our new definition of "identification," we might now turn to the question of what it means to identify harmony *holistically*. I have already discussed two relevant concepts in Chapter 2.4.1: *Gestalt* and *emergence*. As was established, following the definition of Pomerantz and Cragin (2015), these concepts are fundamentally inseparable: When two or more objects group perceptually, salient and novel features emerge from their juxtaposition in time or space. Such emergent features make the whole—the Gestalt—qualitatively different from the sum of its parts. In addition, I make use of the term *Gestalt quality* to refer to the experiential dimension[41] of perceiving a Gestalt. Before returning to these concepts, however, let us review a couple of discussions of "holistic" harmonic listening found in the aural training literature.

Karpinski himself, in a later work, offers an apparent definition: "When you are able to identify something simply because you know what it is, not because you have analyzed or manipulated it, then you have identified it holistically" (Karpinski, 2007, p. 157). On closer inspection, this is not a definition as much as it is a description. It mainly approaches the meaning of holistic identification through negation: Holistic identification does not involve analysis or manipulation. We are still in the dark concerning what it *does* in fact involve, more than *simply knowing what we hear*. Later in the same book, however, Karpinski is more specific. In a section on seventh chords, he recommends practicing the material in two ways:

- Sing the pitches that make up the seventh chord and identify the component intervals;

41 Cf. *qualia*, or *effect* (Doll, 2017).

- Listen to the overall effect and affect of each seventh chord and learn to recognize each seventh-chord quality as a whole. (Karpinski, 2007, p. 202)

The second approach contains several noteworthy words: "Overall effect" seems identical to *Gestalt quality*, i.e., the experiential dimension of hearing a seventh chord. The same goes for "seventh-chord quality," which should be learned to "recognize … as a whole." Finally, the "affect of each seventh chord" hints at an emergent feature which is different in each of type of seventh chord. These are all holistic features, i.e., qualities that pertain to the totality of the seventh chord(s), and not to any of their constituent parts.

Rahn and McKay (1988) offer another definition of *holistic*. They conceptualize possible techniques for harmonic dictation as strategies along a continuum. At the one extreme of this continuum are methods they term *reductionist*:

> According to this approach, students might arrive at a conclusion about the whole (a chord or a chord progression) on the basis of its smallest, most detailed parts (the individual notes of the chord or the individual lines of the texture). (Rahn & McKay, 1988, p. 101).

The drawback of such methods, according to the authors, is a too nearsighted approach to harmonic listening, whereby a small mistake in one of the parts may cause a great error in the assessment of the overall harmony. I would add that with a reductionist listening approach, one may in fact not be listening to the *harmony* at all, but rather to its atoms (e.g., individual pitches, lines, or intervals), making it a harmonic listening strategy of limited value.

At the other extreme of the continuum, we find the *holistic* listening method, whereby one "might be encouraged to leap directly to a conclusion concerning a chord or chord progression merely on the basis of its overall sound" (Rahn & McKay, 1988, p. 101f). We notice again the use of a Gestalt-like descriptor, "overall sound." Like in Karpinski's example, this transfers to what I call the chord's Gestalt quality. It pertains to the totality of the chord as a perceptual object, rather than to any of its constituent parts. In this regard, Rahn and McKay seem to be aligned with Karpinski in what distinguishes a "holistic" form of harmonic listening.

However, the wording in Rahn and McKay's description ("leap … to a conclusion") gives away that the authors do not consider the holistic listening method a good approach to harmonic dictation. They explain why: It's a "hit-and-miss" strategy, where students might misinterpret a chord completely. What Rahn and McKay are in fact problematizing, however, is that students may lack the *ability* to identify the chord or chord progression holistically. If identification

is the ability to detect, recognize, and understand relevant properties of the musical percept, and to communicate this understanding, it is obvious that the students in Rahn and McKay's example are not *identifying* the harmony in question. They are instructed to try, by means of its "overall sound." But if they have not developed the necessary perceptual concepts for recognizing and discriminating the harmonic material in such a manner (cf. DeBellis, 2005, p. 56), it is indeed a hit-and-miss strategy.

In other words, Rahn and McKay's objection does not address holistic identification per se, but holistic "quasi-identification," or indeed, *guessing*. "True" holistic identification (i.e., as portrayed by Karpinski) is not guessing, the same way that recognizing one's mother's voice on the telephone is not guessing. Therefore, while Rahn and McKay draw attention to a relevant pedagogical problem with using a holistic approach to harmonic listening, their argument does not pertain to the precision and dexterity of genuine holistic identification skills. Rather, they point to the same obstacle as Karpinski: "Listeners at any given stage of development either recognize a particular chord or don't" (Karpinski, 2000, p. 119). For listeners who don't, holistic identification is not yet available. If aural teachers would like to help them develop this ability, we seem to need a different, more specific approach than the one described by Rahn and McKay. We must distinguish between holistic listening as a *strategy* (cf. Rahn & McKay) and holistic identification as a *skill* (cf. Karpinski).

A similar objection can be directed at Jarvis (2015). With explicit reference to Karpinski's *Gestalt listening*, Jarvis argues that the average aural training student already has certain holistic harmony hearing skills (e.g., the ability to identify the *Let It Be* chord progression). His aim is to demonstrate how teachers can nurture this ability through statistical learning and repeated exposure to common harmonic progressions. Like Rahn and McKay, Jarvis describes an aural training approach whereby students are repeatedly asked to listen holistically and without active analysis: "The real-time nature of the activity allows Gestalt listening to be developed by forcing students to decide on a chord's function quickly without relying on other means like retrospective harmonic analysis of multiple melodic lines" (Jarvis, 2015, para. 12). As with Rahn and McKay, one might also in this case object to the missing distinction between holistic listening as a strategy and holistic identification as a skill. In this sense, Jarvis's listening strategy *mimics* holistic identification: It is a kind of "fake it till you become it" strategy.

However, there is one important difference between the strategies described by Jarvis and Rahn/McKay. Jarvis's method prescribes that whenever a student misidentifies a chord in a progression, the instructor should help them reevaluate their error. By playing back both the correct version and the version suggested by the student, the instructor lets the student aurally compare and explore the difference between the two versions. This, in my opinion, is

a significant improvement of the "hit-and-miss" procedure discarded by Rahn and McKay, and apparently the main reason why Jarvis does not share their conclusion on the efficacy of a holistic listening approach. Better yet would be the introduction of clear metacognitive strategies[42] on the students' behalf beyond reevaluating perceptual mistakes. Jarvis's method is still fundamentally based on the same principle as that put forward by Karpinski (2000, p. 119)—repeated exposure, and the assumption that brains will eventually have "absorbed [the harmonic material's] regularities through some process of statistical learning allowing them to recognize it with little-to-no conscious effort" (Jarvis, 2015, para. 3).

Partly looking to the aforementioned examples from the aural training literature, but mainly leaning on the concepts of *Gestalt* and *emergence* as presented in this dissertation, I will offer the following definition of "holistic" in the context of harmonic listening:

Holistic listening is to direct one's attention towards a chord or chord progression's overall quality—i.e., its emergent features—rather than towards any of its constituent parts.

Note that this definition concerns holistic *listening* in a general sense, not holistic *identification*. It specifies a perceptual approach, not the efficiency, accuracy, or outcome of that approach. This is useful because it lets us concretize the contents of the "holistic listening" approaches portrayed by Jarvis and Rahn and McKay respectively. Holistic listening, in this sense, is the opposite of an "atomistic" or "reductionist" listening approach, regardless of whether it leads to identification or not. A holistic approach attends to the whole, a reductionist approach attends to one or more of the parts.

The definition offered above may seem to state the obvious, or to simply reformulate what has already been established by, e.g., Jarvis and Rahn and McKay. However, it is worth pointing out that there is not complete agreement in the aural training literature on this notion of "holistic." For example, Stevens (2016) defines holistic hearing as "a synthesizing mode of hearing in which each individual element is considered in relation to the others to form a perception of the whole" (Stevens, 2016, p. 113), and states that using his Do/Ti Test allows students to "identify harmonies holistically based on the presence of particular guide-tones" (Stevens, 2016, abstract). These descriptions both portray the use of a chord's constituent parts to arrive at a judgement of its overall identity, and they would not be termed "holistic" following the definition offered here.

42 *Metacognitive strategies*: "Strategies that concern the way the learner reflects upon their own learning and thought processes" (Blix, 2013).

We may now contrast the definition of holistic listening with a narrower one, which also incorporates the element of identification:

Holistic identification is to detect, recognize, and differentiate relevant properties of a musical percept ***by means of its overall quality—i.e., its emergent features****—and to be able to express these properties through a corresponding, external response.*

There is one criterion missing in this definition, a criterion that is something of a recurring theme in the discussions of "Gestalt listening": The lack of analysis or manipulation (Karpinski, 2007, p. 157), "conscious effort" (Jarvis, 2015, para. 1), "intellectual scrutiny" (Karpinski, 2000, p. 119), etc. While references to a "non-analytical" form of harmony identification are not uncommon in the aural training literature (as also shown earlier in this dissertation), I have chosen to exclude this aspect from the definition. The reason is that it does not seem viable to operationalize the term "analysis." What counts as "analysis" in a perceptual act? Where do we draw the line between analytical and non-analytical components of perception? How would we assess whether such components are involved or not in a particular act of identification? In this regard, the term "analysis" is no more useful than, say, "effort." How would we determine whether an identification of a chord is "effortless" or not?

Furthermore, if "analysis" refers to *atomistic methods*—for example, listening for particular chord tones, the bass line, etc.—then this is already covered by our definition: Holistic listening is directing one's attention towards the overall quality rather than towards any of the constituent parts.[43] If, on the other hand, "analysis" pertains to the *overall quality* of the chord—for example, the momentary rummaging through long-term memory for the correct music-theoretical label to match the Gestalt quality—then it is not obvious that this should disqualify the said identification from being "holistic," and why. Finally, if "non-analytical" is meant to signify the perceptual *speed* with which something is identified—cf. "spontaneous," "intuitive," "directly," etc.—this is a more valid consideration. I will argue, however, that this would create an artificial and purposeless boundary between rapid and slightly-less-rapid identifications. Therefore, I argue that the definition of holistic identification should limit itself to *perceptual focus* (i.e., holistic properties) and *outcome* (i.e., correct perceptual interpretation). This captures the element of perceptual *precision*, from which perceptual speed will follow.

43 Although an *awareness* of the constituent parts may be encompassed in holistic perception: Holistic identification of a Phrygian cadence includes an awareness of the associated bass motion.

4.4 What is the meaning of *harmony*?

Musical harmony is an immense and complex subject, both as an object of perception and as a theoretical system. It is necessary to clarify what is meant by "harmony" in the context of holistic harmony identification, without encompassing the totality of this complexity.

Grove Music Online offers a general and, at first sight, quite straightforward definition of harmony: "The combining of notes simultaneously, to produce chords, and successively, to produce chord progressions" (Cohn et al., 2001, para. 1). This definition points to two important dimensions of harmony, both of which concern its temporal aspect: *Simultaneity* and *succession*.

Before I move on, a short comment on nomenclature: In Western musicology, simultaneity and succession are often referred to as the "vertical" and "horizontal" dimensions of harmony respectively.[44] This is a metaphorical conceptualization of pitch-space which is also expressed in our musical notation system: Frequency (i.e., pitch height) is distributed along a vertical axis, while duration follows the horizontal axis. In general, conceptual metaphors[45] like these highlight selected properties of a domain while obscuring others (Lakoff & Johnson, 1980, p. 10). In the case of our Western musical notation system, the vertical/horizontal pitch-space conceptualization allows us to specify frequencies, rhythms, and durations with great precision. On the other hand, this notation system has no means by which to distinguish between the timbral qualities of 440 Hz performed on a clarinet, and 440 Hz performed on a chainsaw—timbre is obscured by the verticality/horizontality metaphor. Incidentally, it is well established that this is only one of several possible conceptualizations of pitch used in cultures and languages around the world (e.g., Antovic, 2009; Dolscheid et al., 2013; Zbikowski, 2002). For reasons like these, I will refrain from using the ubiquitous terms "vertical" and "horizontal" about harmony in this study. I will instead adhere to those used above by Cohn et al.—the *simultaneous* and *successive* dimensions of harmony.

The simultaneous and successive dimensions of harmony imply several notions of harmonic grouping. Therefore, we must elucidate these to consider their relevance for holistic harmony identification.

First, there is the notion of grouping perspicuously expressed in Cohn et al.'s definition: "The combining of notes simultaneously, to produce chords." The grouping of simultaneously

44 For example, by Radocy and Boyle: "The horizontal dimension encompasses pitch sequences. The vertical dimension involves simultaneous pitch structures" (Radocy & Boyle, 2012, p. 228).

45 Conceptual metaphors, in the tradition of Lakoff and Johnson (1980), are discussed in greater detail in Part II of this thesis.

sounded pitches into chords is uncontroversial. It is a fundamental principle both of psychoacoustics (cf. Bregman, 1990, p. 496) and of modern harmony as a theoretical system: A slow evolution between the sixteenth and eighteenth centuries gradually moved away from viewing harmony as an end-product of counterpoint, towards regarding the chord a primary element (Cohn et al., 2001, 2.[i]). This basic grouping is—at least in principle—largely independent from the successive aspects of harmony. Nonetheless, the "meaning" of a single chord is both perceptually and theoretically influenced by the larger context in which it appears. The *Tristan* chord, from the Prelude to Richard Wagner's music drama *Tristan und Isolde*, is an obvious and extensively debated example of this (see e.g., Nattiez, 1990, chapter 9 for a review).

The next part of the definition is somewhat more problematic. "The combining of notes [sic] … successively, to produce chord progressions" has a double meaning. On the one hand there is the combination of successive *chords* into chord progressions, cadences, etc. This notion of grouping is also uncontroversial, as it expresses the conventional understanding of the term *chord progression*. On the other hand, however, the definition can be understood the combination of *single pitches* into chord progressions. This notion of grouping is less lucid and demands closer examination. It may refer to the grouping of temporally distributed (e.g., arpeggiated) chord tones. This differs from the grouping of simultaneous pitches in that the pitches have non-identical onsets, and therefore do not form a single acoustic object. However, the pitches may still variously be perceptually grouped due to the principles formulated in the Gestalt laws.[46] Johann Sebastian Bach's *Prelude in C major* (BWV 846a) from *Das wohltemperierte Klavier* may serve as an example: In the whole prelude, we are never presented with a C major simultaneity, in the sense that the pitches C, E, and G are sounded with common onset. Indeed, we are not presented with *any* harmonic simultaneities in this sense until the tonic chord that ends the piece—and even this chord is frequently arpeggiated in performance.[47] Yet, the pitches in bar 1 (see Figure 3) are highly prone to harmonic grouping, for several reasons: Tones are sustained as they enter, thus *sounding* simultaneously even if having non-simultaneous onsets. In addition, the tones form a regular, ascending pattern (cf. *the law of good continuation*). Thus, the contents of bar 1 afford the perceptual quality of a major chord. This "majorness" emerges from the pitches' juxtaposition, just as distinctly as it would from the same pitches sounded with common onset. Bar 1 is furthermore perceptually distinct from bar 2, because bar 2 repeats the ascending pattern with changed pitch content, suggesting a new perceptual Gestalt (cf. *the law of closure*). Therefore, the claim that single pitches can be grouped into chord progressions is valid in the sense just described.

46 The most commonly accepted "Gestalt laws" state that elements are prone to perceptual grouping when close together (i.e., *the law of proximity*), when similar to one another (*the law of similarity*), when forming a closed shape (*the law of closure*), or when perceived to move in the same direction (*the law of good continuation*) (Rock & Palmer, 1990, p. 85).

47 E.g., recordings by Walter Gieseking (Bach, 1990), András Schiff (Bach, 1984), or Ton Koopman (Bach, 1983).

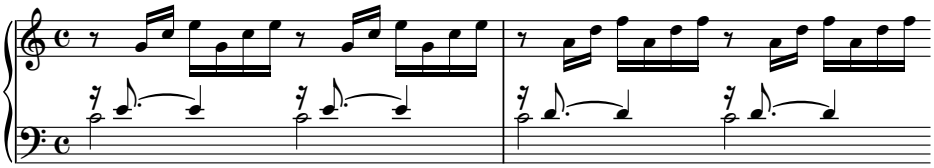


Figure 3: Excerpt from J. S. Bach's Prelude in C major (BWV 846a)

On the other hand, the notion of combining single pitches into chords may refer to *implied harmony*. While simultaneous pitches may be said to constitute *explicit* harmony, implied harmony is the harmony which is suggested by consecutive pitches (Jansen & Povel, 2004, p. 31), usually pertaining to melodies or monophonic music. Implied harmony differs from the *Prelude in C major* discussed above in that pitches are not sustained. Followingly, any harmonic objects implied in such music exist only as a product of the listener's perceptual organization, not as pitches physically sounding together.[48] Implied harmony is an ambiguous concept, however. Music may express strongly implied harmony (e.g., in the *Prelude* from Johann Sebastian Bach's Cello Suite no. 4 in E♭ major [BWV 1010]), more vaguely implied harmony (e.g., in Olivier Messiaen's *Quatuor pour la fin du Temps: VI. Danse de la fureur, pour les sept trompettes* [1941]), hardly any implied harmony (e.g., Luciano Berio's *Sequenza I for Flute* [1958]), or anything in between.

Perceptually, too, implied harmony is a rather elusive phenomenon. While the importance of implied harmony in tonal melody perception has been repeatedly demonstrated (see Kim et al., 2018, p. 595 for a review), much less is known about the perception of implied harmony itself. However, it appears to be relatively complex: In an early study, Povel and Jansen (2002) found that the occurrence of non-chord tones could seriously hamper the harmonic interpretation of tonal melodies. Kim et al.'s more recent study (2018) found that processing of implied harmony was strongly influenced by listener expectations, the placement of chord changes, and even positions of tones within a bar. Furthermore, because implied harmony is a perceptual construct (that is, in a different sense than explicit harmony), we may surmise that it is contingent on yet other factors such as detection of harmonic rhythm, distinguishing between pitch collections, echoic memory, chunking, segmenting, abstraction, etc. The demands on listener capacities involved in these processes arguably places implied harmony

48 Admittedly, the borderline suggested here is theoretical, and may be flexible in real life. For instance, the reverb of a room, the resonating body of a cello, the decay of vibrating guitar strings, articulation, etc. may all contribute to creating pitch simultaneities not present in the notation of a given work of music. The role of echoic memory in this is also rather unclear.

in a different category—both quantitatively and qualitatively—than the types of harmonic grouping discussed above.

Because of the ambiguity and complexity associated with implied harmony, I will keep it outside the scope of this study. This is not because holistic harmony identification does not appear relevant to the perception of implied harmony, or vice versa. On the contrary. But implied harmony is perceptually, music-theoretically, and phenomenologically a rather different concept than harmonic grouping based on simultaneously sounding pitches, and it merits a study of its own.

I started this discussion of "harmony" with the following definition: "The combining of notes simultaneously, to produce chords, and successively, to produce chord progressions" (Cohn et al., 2001, para. 1). This definition had one merit in particular: highlighting and distinguishing the simultaneous and successive dimensions of harmony. An examination of their implications for holistic harmony identification found that the two dimensions comprise several notions of harmonic grouping. Simultaneous pitches form chords, successive chords form chord progressions. In addition, temporally distributed (but simultaneously sounding) pitches may also form chords and chord progressions, whereas temporally distributed (but *not* simultaneously sounding) pitches may form implied harmony. The latter category of harmonic grouping is kept outside this study because of its complexity and dissimilarity to the other categories.

The understanding of "harmony" we are left with, then, can be summed up as "simultaneously attacked or perceptually fused pitches" (Cohn et al., 2001, 5., para. 3). The term "fused pitches" is a slightly unfortunate choice of words, however, because it bears linguistic resemblance to *harmonic fusion* (also referred to as *tonal fusion*).[49] This meaning is not the one intended here—that is, the "fused pitches" do not blend perceptually into a single pitch. Rather, the term refers to basic perceptual grouping, i.e., pitches forming a harmonic Gestalt. Therefore, I propose that a better choice of words is "simultaneously attacked or perceptually grouped pitches."

Our first definition of harmony emphasized its constructional aspect ("the *combining* of notes … to *produce* chords [and] chord progressions" [emphases added]). The second definition complements the first with a more psychoacoustic perspective ("perceptually grouped"). We may incorporate both perspectives in a final definition:

In holistic harmony identification, "harmony" is understood as pitch simultaneities or pitch successions which afford perceptual grouping into chords or chord progressions.

49 See footnote 8 for a discussion of these terms.

4.5 "Holistic harmony identification" as a composite term: Definition and example

Now that I have provided a definition for each of the terms "holistic," "harmony," and "identification" respectively, these definitions may be combined into an overarching definition of the concept of holistic harmony identification:

Holistic harmony identification is to detect, recognize, and differentiate relevant harmonic properties of a musical percept by means of its overall quality—i.e., its emergent features—and to be able to express these properties through a corresponding, external response. Harmonic properties in this regard are pitch simultaneities or successions which afford perceptual grouping into chords or chord progressions.

To assess the precision and functionality of this definition, we should put it to practical use. I will do this using the first chord of the main theme of Mendelssohn's *Wedding March* from *A Midsummer Night's Dream* (op. 61) as an example (see Figure 4). This chord, in this specific voicing, forms a characteristic Gestalt which one might expect to be recognizable for many listeners familiar with Western culture—especially when performed on an organ. Let us imagine what a typical organ rendition of this chord affords for some hypothetical listeners. A non-musician familiar with Hollywood's musical tropes recognizes it as *wedding-y*, without being able to specify exactly why. A wedding coordinator immediately recognizes it as "the Mendelssohn" (as opposed to "the Wagner"). So does the organist of the local church, but in addition, he experiences the motor memory of actually playing the chord: the right hand shaped slightly uncomfortably, the little finger reaching for the high C while the middle finger plays the F♯. A jazz guitarist recognizes the wedding character, but also recognizes the sonority as "minor 6th chord."

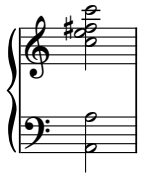


Figure 4: Mendelssohn's "Wedding" chord

As pointed out earlier, Karpinski uses the term "recognize" (or variants of it) eight times in his description of Gestalt listening. The examples given above illustrate the important difference between simply *recognizing* a chord and *identifying* it. It is clear that all these imaginary

persons recognize the chord in question. It is furthermore clear that they recognize it in a holistic manner, since they are all attuned to the cultural affordance of "wedding" (or, specifically, "Mendelssohn"). This affordance is an emergent feature, not something that can be traced back to the constituent parts of the chord. Yet in spite of the "holistic recognition" performed by all these listeners, it is obvious that not all of them demonstrate the expert skill that Karpinski is concerned with. I will examine them one by one.

The non-musician does not perform an identification of the chord, in the meaning we have established. He provides no external response that communicates an understanding of the chord's harmonic properties. He is, however, able to perceive and articulate an emergent feature of the harmonic Gestalt: *wedding-y*. This is what Philip Tagg (2013) refers to as an *esthesic*[50] *descriptor*: it verbalizes an aspect of what the chord sounds like, or what it may communicate to a listener. As I will propose later, such descriptors—albeit crude associations—may not be without merit in an aural skills context. They represent a bridge between (in DeBellis's terms) not having a perceptual concept for a harmonic Gestalt, and having one.

The wedding coordinator is able to identify the piece of music that the chord is associated with. Thus, her external response contains more specific information than that of the non-musician. However, this information does not pertain to relevant harmonic properties of the chord. It merely consists of an association between the chord and a piece of music. Therefore, neither the wedding coordinator performs an identification of the chord.

The organist hears the chord, recognizes which piece of music it represents, and associates it to the experience of performing the piece himself. The motor memory of producing the chord demonstrates that the organist would probably be able to *play back* the chord on an instrument. Correctly playing back the harmonic material is an accepted form of response according to our definition of identification, so this constitutes an identification. Because the identification is achieved by means of the chord's overall qualities (the emergent features of "the Mendelssohn" and motor memory) rather than by atomistic analysis, we may conclude that this qualifies as a holistic harmony identification.

The jazz guitarist also has the wedding association, but simultaneously perceives the chord's Gestalt quality and is able to label this quality verbally. The response is thus given in the form of a music-theoretical term rather than through sonic reproduction. Otherwise, this constitutes a holistic harmony identification for the same reasons as for the organist.

50 *Esthesic*: Semiological term from Molino (1975) via Nattiez (e.g., 1990). Relating to the perceptional aspects of music rather than its constructional or productional (cf. *poïetic*) aspects.

4.6 Summary

The hypothetical examples discussed above demonstrate that our definition of holistic harmony identification is functional. It allows us, with acceptable precision, to distinguish between cases and non-cases of holistic harmony identification, and to specify the justifications for inclusion or non-inclusion in this category.

I have now renamed Karpinski's concept. The main motivation for doing this was the use of two specific terms in Karpinski's presentation (2000, p. 119): "Gestalt" and "recognition." I have made the claim that these terms are problematic. The "Gestalt" in *Gestalt listening* is problematic because it does not pertain to basic perceptual grouping, as in the commonly accepted meaning of the word *Gestalt*. Therefore, its meaning in this context is somewhat ambiguous. "Recognition" is problematic because the skill Karpinski portrays goes further than mere recognition—both in a colloquial and more formal meaning of the word. Consequently, it is not easy to distinguish between the types of recognition that characterize Gestalt listening, and other, non-relevant types of recognition. I have proposed substituting these terms with "holistic" and "identification" respectively—terms that Karpinski himself uses elsewhere to recount the phenomenon referred to as Gestalt listening.

I have explicated the terms "holistic" and "identification" respectively, as well as what is included in the notion of "harmony." I have then proposed the term *holistic harmony identification* with the following definition:

Holistic harmony identification is to detect, recognize, and differentiate relevant harmonic properties of a musical percept by means of its overall quality—i.e., its emergent features—and to be able to express these properties through a corresponding, external response. Harmonic properties in this regard are pitch simultaneities or successions which afford perceptual grouping into chords or chord progressions.

This term, I claim, is better suited at pinpointing the exact meaning of Karpinski's concept than the term *Gestalt listening*. It is also more functional since it specifies the criteria of inclusion. Thus, I propose that the definition I have given for *holistic harmony identification* is in fact valid for the concept Karpinski calls *Gestalt listening*.

5 Holistic listening in the aural training classroom: An ecological exploration

5.1 Introduction

The previous chapter sought to define holistic harmony identification by drawing upon insights primarily from psychological, pedagogical, and aural training literature. This chapter aims at exploring the concept further by using a more explicitly ecological approach. The exploration will have several components. I will assess the aural training classroom as a perceptual environment, reflecting on how this somewhat artificial "habitat" shapes listener intentions and actions. I will discuss musical harmony's ability to specify extra-musical domains, and I will propose that such extra-musical connotations may form perceptual "meanings" relevant to a holistic listening approach. I will use the Neapolitan chord as a case study for investigating various facets of holistic harmony identification, including the role of perceptual learning. Finally, I will examine the relationship between variant and invariant properties of musical harmony. For example, how is it possible to perceive two Neapolitan chords as being the "same," even if they share virtually no physical attributes in terms of timbre, texture, pitch register, etc.?

A thread through this chapter is the investigation of how acts of holistic harmony identification might be experienced from a listener's perspective. The purpose of such discussions is not to present some purportedly objective description of "what it is like" to perceive a given chord or progression—I do not believe such objectivity exists. The subjectivity of sound perception is pointed out, among others, by Nicola Dibben (2001):

> These studies indicate that sounds specify not only physical sources, but cultural and musical attributes as well … These "specifications" are in this sense "meanings" which sounds have for particular listeners. What a listener hears when he or she hears music appears to be "hearing as": materials with meanings guided by listeners' needs and preoccupations. (p. 183)

However, this subjectivity does not imply total idiosyncrasy. The subjectivity is to some degree constrained by the material properties of the object of perception, as well as their social and historical associations (Clarke, 2005, p. 125; Dibben, 2001, p. 185). The reciprocal relationship between these material properties and the capacities and predispositions of an individual perceiver is one of the main tenets of ecological theory. In this way, an ecological approach

transcends the objective/subjective dichotomy by focusing on the classification of experiences rather than assuming an "outside" standing over an "inside" (Reybrouck, 2021, p. 73).

An overall aim of the present chapter, then, is to explore the conditions for the subjectivity of holistic harmony perception: how certain attributes of harmony afford a certain range of possible perceptual interpretations, and how an empirical listener may resonate to these attributes. In this attempt to combine interpretation with a consideration of perceptual principles, I follow the example of Clarke (2005), who describes this method as a "middle way between the outright empiricism of an ethnomethodological approach, and the potential authoritarianism of a hermeneutic approach" (p. 123). Through this approach, I hope to provide some initial insights into the phenomenological nature of holistic harmony identification—or, in other words, to offer some perspectives on holistic harmony identification *from within* rather than from an outside position.

5.2 The aural training classroom as a perceptual environment

Ecological theory holds that perception is the active engagement with one's environment, and that the search for *meaning* is a central part of this engagement (Clarke, 2005). What characterizes the aural training setting as a particular kind of environment, and what "meaning" is there to be found?

First, an aural training setting predisposes students towards particular perceptual goals. The perception has an *intention*, and is often expected to lead to certain type of musical behavior. This intention can be, for instance, the auditive identification of specific musical properties (e.g., a harmonic progression) or error-detection (i.e., comparing a musical score to played-back music with the purpose of identifying intentional notational errors). Generally, intentional listening influences perception by creating a filter that separates the relevant from the irrelevant (Imhof, 2020, p. 235). For example, a harmonic dictation might accentuate an attentiveness to chord quality and bass line, at the expense of other features like lyrics, timbre, or rhythm. Using Thoresen's (2015) classifications of listening intentions (based on the work of Pierre Schaeffer (1977)), we may refer to this as *selective listening*,[51] characterized by the selection of a particular strand of sound, "often seeking to match it against a pre-existing notion or category" (Thoresen, 2015, p. 8). It is furthermore a *specialized* mode of listening, "often determined by professional interest" (p. 10).

51 Schaeffer's term in French is *entendre*.

Second, the aural training setting constrains and specifies the available affordances. Affordances are the possible interactions between the environment and the perceiver. In aural training, the act of perception is generally expected to result in an action which is often largely predefined: the identification of relevant musical structures, and the communication of these through a corresponding, external response. For instance, perceiving a chord progression may afford the translation of these chords into visual symbols, or the reproduction of these chords on a piano. In this regard, the perceptual "meaning" that the aural training student searches for when hearing a chord progression, is the identification of the chord progression's content.

Third, social factors may influence student performance in the aural training classroom, for example performance anxiety (see, for example, Mishra, 2000; Ozturk & Kalyoncu, 2018; Reitan, 2008). While this falls outside the subject of the present dissertation, it is important to acknowledge that the aural training setting in itself can create special obstacles to perception.

In an ecological sense, then, music in an aural training situation generally affords other forms of listening, and other forms of action, than music in other social contexts. However, this somewhat contrived musical setting may still reflect the listening intentions and affordances involved in more "natural" musical interaction, as exemplified by jazz theorist Mark Levine (1995) in the following scene:

> If you change a major chord to Lydian augmented while soloing, the musicians 'comping for you will hopefully hear what you're doing and play $^{\Delta\sharp 5}$ [a major-seventh chord with a raised fifth] also. And if you're 'comping, you'll want to reach the level of expertise where you can hear the soloist doing this and immediately adjust. *Listen!* (Levine, 1995, p. 291)

In mainstream jazz, the common foundation for improvisation is traditionally the tune's harmonic structure. The performance of *I Got Rhythm* at a jam session is primarily a mutual improvisation over the chord progression of that song.[52] This chord progression is not fixed, however, and may be subject to a range of changes, on both single-chord level and globally. The detection of and response to such reharmonizations, both by the soloist and the rhythm section, may thus become an important part of the perceptual meaning to be found in this musical interaction, as described by Levine.

52 Countless jazz standards have been written over the harmonic progression of *I Got Rhythm*, commonly referred to as "rhythm changes."

5.3 Harmony and extra-musical connotations

In the above discussion, perceptual "meaning" is connected to the identification of musical structures. By design, aural training strongly promotes a form of listening that targets such structures. However, if we zoom in on harmonic dictation, there is arguably also another, more subtle level of affordances involved, in which "meaning" can include a more interpretational dimension. This position reflects that of Clarke (2005), who uses a large part of his book to draw attention to "what else *apart* from structure is specified by the sounds of music" (p. 134). What, then, may musical harmony specify, besides harmony itself?

First, a note on the term "extra-musical," which is somewhat contentious. In particular within an ecological approach, the intra-/extra-musical distinction may be viewed as a false dichotomy (see Clarke, 2005, p. 157), since perception is understood in terms of environment–perceiver interactions. There is really no "outside" to this. What I mean by *extra-musical* in the following discussion is simply that harmony, under certain circumstances, has the capability of specifying other domains than that of music.

Musical harmony frequently carries extra-musical connotations. While it is always intertwined with other parameters in actual musical expression, some such connotations are relatively harmony-specific. One example is the large body of empirical studies describing the association of major and minor with positive and negative emotional valences respectively (i.e., "major is happy, minor is sad").[53] When music elicits connotations like these, they are often referred to as *emotions*, or emotional qualities (e.g., Juslin & Sloboda, 2010; Kastner & Crowder, 1990; Lahdelma & Eerola, 2016; Meyer, 1956). However, Jackendoff and Lerdahl (2006) argue that *affect* is a better term, because it allows for a broader enquiry of connotations:

> We do not deny that, say, *Yellow Submarine* is happy and *Michelle* is rather sad, and that these judgments are correlated with their respective modes (major and minor), rhythms, and pitch ranges. But there is a vastly wider range of descriptors that deserve characterization. A passage of music can be gentle, forceful, awkward, abrupt, static, earnest, opening up, shutting down, mysterious, sinister, forthright, noble, reverent, transcendent, tender, ecstatic, sentimental, longing, striving, resolute, depressive, playful, witty, ironic, tense, unsettled, heroic, or wild. Few of these can be characterized as emotion per se. (Jackendoff & Lerdahl, 2006, p. 60)

53 This association has been found frequently enough in research (e.g., more recently by Lahdelma & Eerola, 2016) to prompt claims that it has a biological rather than cultural basis (e.g., Cook & Hayashi, 2008). However, see Athanasopoulos et al. (2021) for a recent repudiation of this claim.

For the argument I am making here, this is an important point. For example, consider this comparison between sus4 chords and ♯11 chords, originally made by an instructor at a jazz improvisation camp, and retold by an informant in my master's thesis on jazz musicians' representations of harmony:

> And then he said that you should think of it like opening a door. The sus4 chord, that's like a closed door. And when you come to ♯11, you kind of open the door. It's a very *open* sound. And everybody got it. Everybody was able to play on that chord. No one missed it. (Langfeldt, 2017, p. 93, my translation)

The "openness" of the ♯11 chord, and the "closedness" of the sus4, are hardly *emotions*. Yet in this particular situation, they were meaningful conceptualizations of the "moods" of these chords. Therefore, I will follow the example of Jackendoff and Lerdahl and use the broader term *affect* here. In any case, it should be stressed that *perceiving* affective qualities (or "emotions") in music is not the same as *experiencing* them as a listener (Gabrielsson & Juslin, 2003, p. 503). Jackendoff and Lerdahl put it well: "Witty, mysterious, or sinister music does not make one feel witty, mysterious or sinister, but rather as if in the presence of someone or something witty, mysterious, or sinister" (Jackendoff & Lerdahl, 2006, pp. 65–66).

While previous interest in the affective connotations of harmony have consisted primarily of theoretical studies (e.g., Cooke, 1959), some empirical studies have been performed. Pallesen et al. (2005) discovered an effect of musical sophistication on the perceived "sadness" in minor chords. By using fMRI, they found that in a passive listening mode, there were no significant differences between musicians' and non-musicians' neural responses to single musical chords. In this passive state, minor chords (and dissonant chords) elicited responses in brain areas associated with an "alarm system", whereas major chords did not. However, when asked to rate chords on an 11-point scale from "sad" to "happy," musicians rated minor chords as significantly more "sad" than did non-musicians (Pallesen et al., 2005, p. 452). The authors suggest that "the group difference in the emotional ratings of the chords, but not in the neural responses, may reflect a musician's ability to recognize and categorize the chords in terms of the conventional emotional connotations" (Pallesen et al., 2005, p. 452). In other words, a trained musician's categorization of "minorness" appears to be learned, overriding the neural response. In a different study, Lahdelma and Eerola (2016)[54] found that single chords were able to convey musical emotions to listeners. In particular, *nostalgia* was found to be effectively communicated by the minor triad, the minor seventh chord, and particularly the major seventh chord. In a subsequent article (Lahdelma & Eerola, 2015), the authors—like

54 The study appears to have been first published online in October 2014, but I am referencing here the version that appeared in *Psychology of Music*, vol. 44, issue 1 in 2016.

Pallesen et al.—consider learning an important part of emotion perception in chords. However, they also consider the role of intrinsic emotional connotations arising from tonal relations, as well as "clashing conventions":

> The "clashing" of two highly conventionalized indices (the major triad as an index of happiness/joy, and the dissonant intervals of the minor second and the major seventh as indices of negative valence and melancholy/sadness) could create new emotional meaning that is perceived as the complex emotion of nostalgia. (Lahdelma & Eerola, 2015, p. 252)

The three explanations proposed by Lahdelma and Eerola seem congruent with the ecological model of subject–object reciprocity: Whether or not musical chords afford affective connotations to an empirical listener depends on the interrelationship between objective features of the chord and the capacities of the listener.

Another category of harmony's extra-musical connotations is that of cultural references. This category can be seen in connection with *topic theory*, the study of conventional musical signs (see, e.g., Mirka, 2014). One example is the "Mendelssohn chord" discussed earlier. Mendelssohn's *Wedding March* is such a firmly established musical trope for the notion of "wedding" in Western popular culture, that the opening chord alone may evoke wedding connotations—or, in ecological terms, afford the interpretation "wedding." Tagg (2013) suggests another chord with strong cultural connotations: the "James Bond chord." Music-theoretically this is a minor chord with an added major seventh and major ninth, consisting of the tones $\hat{1}$, $\flat\hat{3}$, $\hat{5}$, $\natural\hat{7}$, and $\hat{9}$ (see figure 5).

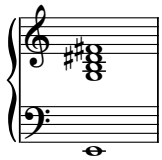


Figure 5: The "James Bond chord"

The chord is a staple of certain musical styles, in particular jazz harmony. To a broader audience, however, it is perhaps most recognizable from its salient appearance at the end of the *James Bond Theme* (written by Monty Norman and arranged by John Barry). By association, it has also become a musical trope for spy or detective movies more generally (perhaps especially when used tongue-in-cheek). Tagg's anecdotal claim is that non-musicians are generally able to aurally distinguish the minr or major ninth chord from other chords if is conceptualized

for them as "spy chord" or "detective chord."[55] This quality, he claims, is still recognizable to many non-musicians when the chord is played on a piano rather than with the reverbed Fender Stratocaster sound used in the original Bond theme. Therefore, the spy connotation is at least partly contained in the chord's tonal information (Tagg, 2013, p. 339)—a point that corresponds to an ecological understanding of cultural affordances.

5.4 Harmony and implicational meaning

What is the significance of extra-musical connotations like those discussed above? First, I contend that they are generally qualities that pertain to the harmonic Gestalt, i.e., *holistic* qualities. They are features that emerge from pitch grouping, and which cannot be traced back to the constituent pitches. The "spy" quality cannot be found in any of the individual tones of the Bond chord. It is only when all these tones are sounded simultaneously that they afford the "spy" interpretation. Therefore, while extra-musical connotations are perhaps "pseudo-conceptualizations" of musical harmony, inferior to music-theoretical concepts, they may in fact assist the holistic recognition of chords.[56] Noticing such connotations is to pay attention to the holistic qualities of the chord in question, and to the affects or associations these holistic qualities might convey. This again may contribute to the "acquisition of a certain repertoire of perceptual concepts for musical events and properties, where one initially has no perceptual concepts of them" (DeBellis, 2005, pp. 55–56). Tagg implies the same when he maintains that non-musicians' ability to recognize or distinguish chords with the aid of esthesic descriptors illustrates that the problem is not necessarily a lack of aural competence (cf. a lack of *perceptual concepts* in DeBellis's terminology), but a lack of poïetic nomenclature (i.e., the theoretical names of chords, cf. *music-theoretical concepts* in DeBellis's terminology; Tagg, 2013, p. 340).

Second, I propose that such connotations may form "meanings" in the perceptual environment of aural training. Rather than mere projections by the listener, they are interpretations afforded by the harmony's materiality. They are not the *only* interpretations available: Clarke points out that music may specify a whole range of objects or events simultaneously, and that any individual listener's attunement to the different interpretational opportunities may vary (2005, pp. 158–159). For example, we could picture a jazz pianist who is so intimately familiar with the use of the minor major ninth chord that it carries no extra-musical connotations:

55 Tagg also describes other "aesthesically labelled sonorities" which are purportedly equally recognizable for non-musicians, for example the "bitter-sweet chord" and the "romantic pathos chord" (Tagg, 2013, p. 340).

56 Note that I use the term *recognition* here, not *identification*. Earlier in the study, recognition was defined as "the automatic activation of some particular contents of long-term memory that have some relation or association with current perception" (Snyder, 2000, p. 10).

it is just a very common voicing of a minor chord. For this pianist, the "spy" connotation is not mobilized.

Still, is it reasonable to depict extra-musical connotations as "meanings"? On the count of affect, Jackendoff and Lerdahl initially seem to disagree: "We do not want to say that affect or emotion is the 'meaning' of music, in the sense that language is meaningful. Unlike language, music does not communicate propositions that can be true or false" (Jackendoff & Lerdahl, 2006, pp. 60–61). There are two objections to this. First, we must distinguish between *musical meaning* (cf. Jackendoff and Lerdahl) and *perceptual meaning*. These are not identical. The former is a hermeneutic construct, while the latter concerns how a listener makes sense of perceptual information available in the sonic environment. In ecological theory, this sense-making is expressed through the concept of affordances—"the ecological equivalent of meaning" (Reybrouck, 2021, p. 72). Therefore, ecological theory has a more inclusive notion of meaning than other theoretical traditions (Clarke, 2017, p. 528). Second, the argument I am making here is not that affective connotations (or other extra-musical connotations) are *the* meaning involved in the sense-making of musical harmony—simply that they are part of the "whole variety of objects or events" that music is capable of specifying (Clarke, 2005, p. 158).

A distinction offered by Barnard (2012) may help to further elucidate the kind of meaning implied here. In order to understand the meanings of music, he suggests, we must distinguish between *propositional* and *implicational* meanings. The former denotes types of meaning that can be clearly expressed. This is the kind of meaning Jackendoff and Lerdahl are referring to when they maintain that "music does not communicate propositions that can be true or false." Implicational meaning, on the other hand, "blends ideas, externally derived percepts and bodily experience … [They] are more abstract schemata or models of deep multimodal interdependences that equate more with latent senses of meaning or knowing as well as feelings, intuitions and affect" (Barnard, 2012, p. 71). While the two strands of meaning coexist and depend upon each other, musical expertise and aesthetic appreciation are more rooted in implicational meanings (p. 71). Because implicational meaning is more easily experienced than expressed in words, it has a certain ineffable quality. Therefore, Barnard notes that implicational meanings are generally more accurately captured through poetry, metaphor, and parable (p. 78).

Adapting Barnard's argument to the context of harmonic perception, I propose that harmony's extra-musical connotations can be seen precisely as *metaphorical* qualities. By metaphorical, I simply mean what is implied in the term "extra-musical," namely that harmony may be experienced "as"—or as expressive of—something that lies outside the domain of music.

In the discussion above, I have barely touched upon such cross-domain associations[57] in musical harmony. I have used a couple of rather salient examples—the "Mendelssohn chord" and the "spy chord"—to illustrate how extra-musical connotations may be understood as *affordances* in an ecological perspective, and as part of the perceptual meaning available in a sonic environment. Furthermore, I have made the argument that they may aid the sense-making involved in harmonic dictation. In Part II of this dissertation, I will conduct a more systematic study on cross-domain mapping of harmony, by analyzing the metaphorical language of 20 textbooks on harmony. In the remainder of Part I, however, I will focus on further exploring the relevance of "metaphorical" listening for the development of holistic harmony identification skills.

5.5 "Subdominantness" and "Neapolitan-ness"

For the first discussion, I will examine Karpinski's example of recognizing a subdominant ("to recognize the 'subdominantness' of the subdominant chord," Karpinski, 2000, p. 119). The problem with Karpinski's example, as pointed out by Chenette et al., is that the subdominant chord is "arguably not a single object but a collection of objects that differ in timbre, texture, inversion, spacing, function [sic], and placement within a phrase" (2021, p. 149).

There are several things to say about this. First, Karpinski does in fact give some context in a footnote as to what kind of subdominant function he has in mind, as well as its placement in a phrase:

> For example, in introducing the subdominant, Trubitt and Hines (1979) write: "It is this triad that is clearly heard in the IV–I (plagal) cadence at the end of numerous hymns, the 'Amen.' Aurally, the I–IV or IV–I progression has a mellower, less intense effect when compared to the I–V or V–I" (p. 104). This is roughly equivalent to Rahn and McKay's (1988) "holistic" approach (pp. 101–2). Also see Rogers's (1984) attention to "the affective and psychological response of certain patterns" (p. 122). (Karpinski, 2000, p. 119, footnote)

The reference to the "Amen cadence" (i.e., a plagal cadence in major) indicates that Karpinski is specifically referring to a major subdominant chord in a major key, in a relatively specific

57 I use the term "cross-domain associations" in this context rather than the more established "cross-domain mapping," because the latter usually implies a mapping of *structures* from a source domain onto a target domain (e.g., Zbikowski, 2002, p. 13–14). The "Mendelssohn chord" and the "spy chord" are more superficial associations, and not expressive of larger, underlying metaphor structures.

context. Thus, his example is not so vague as to be far-fetched. As for Chenette et al.'s concerns about the differences in timbre, texture, inversion, and spacing, I will return to these shortly in a discussion on the principle of invariance.

Second, even if no context were given beyond the relatively unspecific "subdominant," there is no clear reason why a subdominant function should not be considered a relevant harmonic property for holistic harmony identification. It is less specific than for example "major subdominant chord with added sixth." However, one can easily imagine a harmonic dictation exercise that instructs students to identify the individual chords of different progressions as either *tonic*, *subdominant*, or *dominant* respectively. If a subdominant function is then identified holistically, it is a valid case of holistic harmony identification according to the definition proposed in the previous chapter. In fact, in this regard, the term "subdominant" is no more problematic than, say, "red." "Red" is a broad color category that contains a multitude of specific shades, e.g., scarlet, vermilion, burgundy, carmine, ruby, etc. Certain situations require us to pinpoint an exact shade, for example when ordering paint. In other situations, the broader term "red" is perfectly appropriate (e.g., when describing the color of a pill). We might simultaneously perceive a shade of lipstick to be "red" and "maroon"—these are just different levels of specificity that do not cancel each other out. Likewise, there is no inherent contradiction in perceiving a chord simultaneously as "Am^6" and as "subdominant." This is indeed possible with the chord from Mendelssohn's *Wedding March* discussed in the previous chapter. The fanfare that precedes the theme suggests a C major tonality (which is, eventually, established as the main key). The first three chords of the theme, however, are a tonicization of C major's III, E minor (Am^6, B^7, Em). In a slightly larger harmonic context, it is therefore both appropriate and relevant to identify the Am^6 chord as "applied subdominant," even if this label does not specify the exact structure and quality of the chord.

Still, Chenette et al. make a valid point: Karpinski's example would profit from using a more specific chord. Therefore, I will use a more specific chord type as my example in the following discussion: the Neapolitan chord. A Neapolitan is a chromatic predominant[58] chord built from the flattened second degree of the scale, typically (but not necessarily) in a minor key. Though the structure and use of the Neapolitan function was extended in Romantic harmony, the first-inversion triad is the prototypical form frequently found in Baroque and Classical harmony—hence the standard term *Neapolitan sixth chord* (Drabkin, 2001). A typical use is shown in figure 6.

58 Although in late nineteenth-century harmony, Neapolitans lead as frequently to the tonic as to the dominant (Harrison, 1994, p. 116).

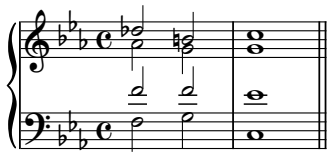


Figure 6: Cadence with Neapolitan sixth chord

What perceptual interpretations do Neapolitan chords afford? If we concentrate on esthesic descriptions—how Neapolitans allegedly *sound*—there is a clear tendency to be found in harmony textbooks (all emphases added):

> "Mozart has used a Neapolitan sixth … as a **colorful** variant." (Clendinning & Marvin, 2016, p. 552)
>
> "**colorful** Neapolitan harmonies" (Clendinning & Marvin, 2016, p. 553)
>
> "special harmonic **coloration**" (Laitz, 2016, p. 562)
>
> "One of the more **colorful** chords that can be used to precede the dominant is the Neapolitan." (Kostka et al., 2018, p. 364)
>
> "its aural effect is **strikingly different** … become familiar with the **distinctive sound** of the [Neapolitan]." (Kostka et al., 2018, p. 365)
>
> "the **color** of the Neapolitan sixth is recognizable." (Piston, 1987, p. 410)

Neapolitans are often highlighted as a *colorful* harmonic function. While we may pass over the exact meaning of this metaphor for the time being,[59] it seems to suggest a certain aural distinctiveness or perceptual salience. This is supported by Lovell (2021), who in a discussion of college-level music theory curricula reports the results of a student assessment. In a section that featured the aural identification of chromatic predominants in an isolated context, 41 out of 48 (85%) sophomore students recognized the Neapolitan chord. Lovell partly attributes the strong result to recognition of the chord's "distinctiveness" (Lovell, 2021, p. 95).

Which properties of Neapolitan chords would explain their alleged distinctiveness? An examination of their tonal information might give a clue. When it appears in a minor key, a

59 The conceptual metaphor HARMONY HAS COLOR is treated in part II of this dissertation.

Neapolitan chord contains one chromatically altered[60] tone (scale degree $\flat\hat{2}$). If appearing in a major key, it contains two (scale degrees $\flat\hat{2}$ and $\flat\hat{6}$). This arguably makes it perceptually more distinct than diatonic chords: its chromaticism makes it "stand out" from the surrounding tonality. This view is supported by Brower (2000), who likens modal mixture to the technique of *chiaroscuro* (a dramatic juxtaposition of light and dark areas) in visual art. The Neapolitan sixth chord, Brower suggests, creates a similar effect by connoting "the positive affect of major in a context in which the chord itself—built upon a root outside the key—is tonally unstable. Perhaps reflecting this, nineteenth-century composers often used the Neapolitan to portray transient or illusory states of happiness" (p. 342). I contend that this Neapolitan distinctiveness—let us call it "Neapolitan-ness," following Karpinski's example—is indeed a Gestalt quality: It is a feature which emerges from the perceptual grouping of certain pitches.

The Neapolitan-ness is not fully explained by the chromatic tone(s) alone, however. Played in isolation, without the context of a tonic, the Neapolitan sixth chord is simply a major chord in first inversion. Rather than possessing the distinct Neapolitan-ness, it now possesses the more low-level Gestalt quality which emerges from the basic perceptual grouping of pitches into a major triad—*majorness*. Neapolitan-ness is a higher-level Gestalt; a feature which emerges from *two* factors: the basic perceptual grouping of pitches into a major triad, *and* the harmonic relationship of this triad to a specific tonic. It is this specific harmonic relationship to its surroundings, rather than its inherent pitch structure, that makes a Neapolitan chord a Neapolitan. This is elegantly demonstrated by the deceptive opening of Camille Saint-Saëns's "Organ Symphony" (*Symphony No. 3* in C minor, op. 78), which begins with a sustained major triad in first inversion. Since no harmonic context is given, the chord affords a vaguely tonic character. Only at the entrance of the second chord does the opening chord change harmonic "meaning" and no longer affords an interpretation as tonic. It retrospectively reveals itself as a Neapolitan chord.[61]

We acknowledge, then, that the precursors for the Gestalt "Neapolitan-ness" are objectively present in the acoustical stimulus—as both Gestalt theory and ecological theory would contend. Thus, the degree of subjectivity involved in Neapolitan perception is not a consequence of ambiguous acoustical information, but of how an individual listener resonates to this information. For example, which associations and esthesic interpretations the stimulus affords—or, in DeNora's (2000) term (via Dibben, 2001), which particular meanings are *mobilized* (or not mobilized).

60 A tone which does not belong to the tonic scale.

61 I am grateful to my colleague Haakon Støring for pointing out this example to me.

5.6 The Neapolitan and perceptual learning

It is worth emphasizing that *perceiving* Neapolitan-ness is not the same as *identifying* it. In the previous chapter I defined identification as detecting, recognizing, and differentiating relevant properties of a musical percept, and being able to express these properties through a corresponding, external response. Since the aural distinctiveness of Neapolitan-ness has a material underpinning, it may in principle be *detected* by any listener, regardless of musical expertise. What does this mean? Simply that it is possible to experience the particular harmonic "color" of a Neapolitan chord, even if it is experienced in a non-conceptual[62] way. After all, non-musicians arguably experience harmony in this manner much of the time they listen to music. Whether or not one is able to *identify* this Neapolitan chord, on the other hand, depends on the level of one's sensitivity towards this particular perceptual information. This includes having a perceptual concept of Neapolitans, possibly also a music-theoretical concept. To translate this to the context of aural training: When a student is not able to distinguish aurally between cadences containing, e.g., an N^6 chord and a $ii^{\circ 6}$ chord respectively (see figure 7) while another student is, the problem does not pertain to lacking perceptual information. The structural properties of those chords contain all the necessary information. Rather, it is a problem of insufficient differentiation and discrimination capacities on the student's behalf. In ecological theory, such insufficiencies are amended through perceptual learning. Again, perceptual learning in an ecological perspective refers to the ability to detect features in the stimulus information that were previously undetected (Clarke, 2005, p. 24), or the increasing ability to distinguish and differentiate between such features (Gibson, 1969; Gibson & Gibson, 1955).

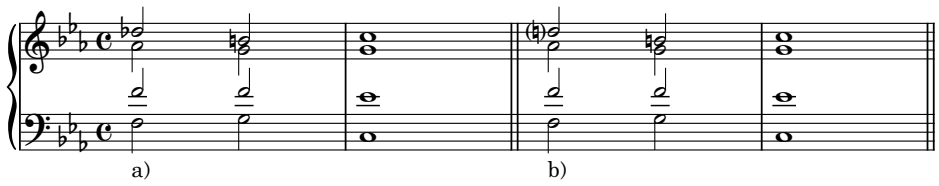


Figure 7: Cadences with a Neapolitan (a) and a $ii^{\circ 6}$ chord (b) respectively

So, how can a teacher aid such perceptual learning? If a student is struggling to aurally distinguish Neapolitans from similar harmonic functions, one might start by diagnosing the problem. On the one hand, it may be caused by the lack of a perceptual concept of Neapolitans. In this case, the student lacks the ability to distinguish perceptually between Neapolitans and

62 Perceptual concepts (cf. DeBellis, 2005) are a function of what a person can recognize and discriminate.

non-Neapolitans, and to perceive different instances of Neapolitan chords as being "the same" (cf. DeBellis, 2005, p. 55). If the music-theoretical concept of Neapolitans is understood (i.e., their structure, tonal content, and harmonic use), only the aspect of perceptual discrimination need be addressed. Several approaches are possible. The student might be encouraged to listen for the characteristic scale degree that often signals a Neapolitan chord: $\flat\hat{2}$. The presence or absence of this scale degree may then be used for deducting deduct logically whether the chord is a Neapolitan or not. This would be an atomistic or "reductionist" approach, following Rahn and McKay's aforementioned description: to "arrive at a conclusion about the whole (a chord or a chord progression) on the basis of its smallest, most detailed parts (the individual notes of the chord or the individual lines of the texture)" (Rahn & McKay, 1988, p. 101). However, in the case of Neapolitans, I contend that an atomistic approach is misguided—or, at the very least, a lost opportunity. The pedagogical advantage of Neapolitans is precisely their aural distinctiveness, which is a *holistic* feature. Thus, effort should be made towards learning to recognize the Neapolitan qua Gestalt, and to distinguish it from structurally similar but esthesically dissimilar chords. In this regard, the statistical learning approach championed by Jarvis (2015) arguably has merit. By continuously contrasting Neapolitans with similar chords in holistic listening exercises, the repeated aural exposure might aid the student's capacity for perceptual distinction. However, one is faced with the challenge pointed out by Karpinski: "Listeners at any given stage of development either recognize a particular chord or don't. In the absence of other listening strategies, they can be taught no concrete means for recognizing a 'new' chord as a Gestalt" (Karpinski, 2000, p. 119). I propose such means exist. First, a further development of Jarvis's statistical learning approach is to analyze the differences in holistic quality between chords that the student fails to distinguish. Let us posit that she has problems distinguishing N^6 and $ii^{\circ 6}$ (see figure 7). These two chords are structurally very similar, the only thing separating them being the lowered scale degree $\flat\hat{2}$ in the Neapolitan sixth chord. Perceptually, however, they are markedly dissimilar, because of a significant difference between their emergent features: The N^6 is a consonant chord (a major triad in first inversion), while the $ii^{\circ 6}$ is dissonant: a diminished chord containing a tritone interval. This salient difference in Gestalt quality should be both stated verbally and experienced perceptually by the student. Second, it is possible to verbalize the Gestalt quality of the Neapolitan chord even further. This relates to the previous discussion on extra-musical connotations and "metaphorical" listening: Hearing a Neapolitan is to hear it "as" something. Naturally, this "something" will vary from listener to listener. For example, to me, the Neapolitan might elicit connotations like "bright," "lifted," "smooth," or "soft." While any given listener may or may not agree with these connotations, they are constant *for me*. Thus, for me, a Neapolitan chord affords these descriptions, whereas a $ii^{\circ 6}$ chord does not. The latter rather affords descriptions like "tense," "hard," and "angular," which makes the two chords quite distinct from one another. While the connotations mentioned here might perhaps be accused of being private, random associations,

it is worth pointing out that they correspond to objective acoustic features of the two chords (i.e., consonance and dissonance respectively). The aim here is to gain a perceptual concept for Neapolitans where one does not have one. The more we are able to verbalize its Gestalt quality, the more we reinforce our awareness of what we hear it "as." While this approach offers no guaranteed pathway to success, it is a practical way of guiding our hypothetical student towards a higher aural awareness of Neapolitan-ness, perhaps even to the point where she is able to identify it holistically. This illustrates that, contrary to Karpinski's claim, holistic chord identification training is indeed conceivable. It also shows that Jarvis's suggested method of extensive repetition can be improved by the introduction of a specific listener intention, guiding and heightening the student's perceptual awareness of chords' sonic materiality.

On the other hand, the difficulty in distinguishing Neapolitans from similar harmonic functions may be caused by a deficient *music-theoretical* concept. If this is the case, the student could be confused about the construction or tone content of Neapolitans, or might even be unaware of the concept altogether. A deficient music-theoretical concept could indicate that the student also lacks a perceptual concept of Neapolitans. However, this need not be the case.[63] This should be clarified in order to determine whether both issues must be addressed, or only the music-theoretical understanding.

5.7 Harmonic invariants

I will now return to the criticism raised by Chenette et al. against Karpinski's "subdominant-ness" example: The subdominant chord is "arguably not a single object but a collection of objects that differ in timbre, texture, inversion, spacing, function [sic], and placement within a phrase" (Chenette et al., 2021, p. 149). While the Neapolitan is a somewhat less ambiguous function than "subdominant" in terms of pitch content, many of the issues remain valid. A Neapolitan chord may be performed on different instruments, in different registers, and with different voicings. Its inversion and harmonic placement may also diverge from the archetypical sixth chord succeeded by a dominant chord. In consequence, different specimens of what

63 On a personal note, this reflects my own situation prior to studying music and learning about Neapolitans. I was well familiar—as a listener, not as a performer—with J. S. Bach's *Passacaglia and Fugue in C minor* (BWV 582) for organ. I particularly appreciated the arrival of that fermata-held, ethereal chord right before the end; a sudden brightening of the harmony; a momentary pause in the incessant sixteenth-note motion of the fugue; a question that seems aimed at the sky, or God, or the world; a precursor to Beethoven's *"muß es sein?"* a century later, mercilessly answered by the *"es muß sein!"* of the final cadence. I had figured out aurally that the chord must be a D♭/F, but I had no explanation for why such a chord would appear in the key of C minor. When I eventually learned about Neapolitans, I immediately understood that the otherworldly Bach-chord was a Neapolitan. To this day it remains, for me, the ur-Neapolitan, of which all other Neapolitans are manifestations. Case in point: I appear to have had a perceptual concept of Neapolitans before acquiring a music-theoretical concept of them.

we refer to as Neapolitan chords may in fact have no common physical attributes. How, then, can we justify regarding them as somehow being perceptually similar—or even "the same?" The answer, from an ecological viewpoint, is the principle of invariance.

What Buccella (2021) calls "the problem of perceptual invariance" is to explain the experience of stability in a perceptual object, even when it undergoes changes. She uses the example of watching a door being opened: Even though the door projects differently shaped images to a perceiver during this movement, it is nonetheless experienced as maintaining a stable shape (Buccella, 2021, p. 13883). Yet in perceptual research, it is not self-evident how this experienced stability—or *invariance*—arises, or where it originates.

Ecological theory has a different approach to this problem than many conventional cognitive approaches:[64] Instead of viewing perceptual invariants as internal representations by the perceiver, it considers them objective features of the perceptual environment (Buccella, 2021, p. 13889). Clarke defines them as "relationships between stimulus properties that remain unchanged despite transformations of the stimulus array as a whole" (Clarke, 2005, p. 34). By placing the source of perceptual invariance in the environment rather than in the perceptual system, the ecological approach resists the cognitive tendency to rely on explanations of internal processes, emphasizing instead the objective features of the phenomenon. (Clarke, 2005, p. 36).

While Gibson was most concerned with visual perception, Clarke suggests that music is particularly suited to exemplify invariance. He points to how themes and motives are frequently altered and manipulated in various ways without losing their identity—that is, we may still recognize them as "the same" theme (Clarke, 2005, p. 35). This is reminiscent of Ehrenfels's (1890/1988) example of the transposed melody: What is a melody, if it remains the same even after we transpose it to a different key, thereby changing all of its components? To Ehrenfels, and the Gestalt movement that followed him, the answer was that melody is a Gestalt. It is not the characteristics of each individual pitch that make up a melody. An attempt to define *Three Blind Mice* as 329.63 Hz sounded for 0.6 seconds, 293.66 Hz for 0.6 seconds and 261.63 Hz for 1.2 seconds etc. would be utterly absurd. Rather, it is the relationship between pitches, and the overall pattern that they form, that make up a melody. This perceptual Gestalt remains fundamentally unchanged even if we transpose it a tritone, and whether we hear it performed on a bassoon, a trumpet, or an electric guitar. Within an ecological approach, this stable quality is referred to as an *invariant property* rather than as *Gestalt* or *emergent feature*. However, both traditions conceive of it as an objective feature of the stimulus rather than something

64 Although the ecological approach (and Gibson's view explicitly) is encompassed in one of the two main approaches to the problem of perceptual invariance, described as the *global-structural* approach by Buccella (2021).

which is added or constructed by the perceiver. Once again, we see the affinity between Gestalt theory and an ecological approach.

Does the principle of invariance help us understand harmonic perception, *holistic* harmony perception, or even holistic harmony *identification*? It does allow us to discuss the idea of *harmonic invariants*, and what these may be. Returning to the example of Neapolitan chords, the various accidental characteristics of any given Neapolitan chord no longer poses a challenge to our understanding of Neapolitan-ness. Neapolitan-ness is an *invariant* feature of Neapolitan chords, a quality which is largely independent and unaffected by variation in timbre, keys, registers, voicings, etc. I have already argued, in accordance with an ecological view, that the precursors for Neapolitan-ness are objectively present in the acoustical stimulus. These precursors pertain simultaneously to the Neapolitan chord's intrinsic pitch structure, and more importantly, its specific relation to the harmonic surroundings. Neither of these are fundamentally altered by changes in timbre, register, texture, etc.[65] Therefore, it should come as no surprise that Neapolitan-ness is a relatively stable quality across superficially differing examples of Neapolitan chords. The same can be said about how vastly different renditions of "12-bar blues" all share certain objective features that allow them to be perceived as being a 12-bar blues.

To restate, I am proposing that Neapolitan-ness is a harmonic invariant for all Neapolitan chords. This establishes Neapolitan-ness as an objective feature of such chords. As perceptual information it is thus available to any listener, whether or not one has a perceptual concept for it: It is possible to *experience* the harmonic effect of a Neapolitan chord, even if one does not have the capacity to *recognize* it, much less to *identify* it. As always, there is an interdependence between stimulus properties and listener capacities. We might compare it to listening to a fugue. In a large-scale fugue, the fugue subject might enter dozens of times, in various voices, registers, and keys, and with a number of transformations. This information is objectively present in the sounded music, consolidating the subject as an invariant of the fugue. However, it is not necessarily picked up by all listeners. Indeed, to an uninitiated listener, a fugue may sound like a chaotic piece of music. To a listener attuned to the style and compositional principles of fugues, on the other hand, it is more likely perceived as highly structured. In other words, there is an important aspect of perceptual learning involved in becoming attuned to stylistic—or harmonic—invariants (Clarke, 2005, p. 35).

65 Although extreme situations may hypothetically blur Neapolitan-ness, e.g., the use of a register so deep that chords cannot be perceptually distinguished.

5.8 Summary

In this chapter, I have discussed the aural training classroom as a perceptual environment, and how this particular type of environment may influence perceptual intention and mode of listening. The role of perceptual learning has also been considered.

I have furthermore discussed musical harmony's ability to specify extra-musical domains, for example affective qualities and cultural references. Tagg's (2013) discussion of the James Bond chord is a salient example of this. I contend that such extra-musical connotations are *holistic* in the sense that they emerge from the chord's overall sound, rather than from its constituent pitches. Thus, attunement towards a chord's extra-musical connotations is a form of awareness (by proxy) of its particular Gestalt quality. I therefore propose that a "metaphorical" listening—i.e., awareness of the implicational meanings that harmony may afford—possibly offers a way of practicing holistic recognition of harmony: Learning to recognize the "spy" quality in a minor major ninth chord is simultaneously a way of learning to aurally recognize and distinguish that specific chord type. If supplemented with theoretical insight of the chord (nomenclature, structure, etc.), such holistic recognition may even be developed into holistic *identification*. In other words, "metaphorical" listening could potentially constitute an intentional training strategy for holistic harmony identification. To evaluate this potential, however, further studies on the role of metaphor in harmony perception are needed. Two such studies will be conducted in Part II and III of this thesis respectively.

Finally, this chapter has examined the perceptual basis for holistic harmony identification from an ecological viewpoint, using the Neapolitan chord as a case study. I have shown that the Neapolitan's aural distinctiveness—or "Neapolitan-ness"—can be explained by objective features of the chord and its relationship to the harmonic surroundings. These objective features contribute to a certain constancy in the perceptual quality of Neapolitan chords, even when other features (e.g., timbre, register, inversion, etc.) are radically changed. This constancy is explained by the principle of invariance. The perceptual pick-up of harmonic invariants will necessarily depend on the individual listener's level of expertise. Yet, by regarding invariants objective features of the perceptual environment, rather than internal representations by an individual perceiver, ecological theory offers a cogent explanation of Karpinski's "subdominantness" example, which was challenged by Chenette et al. (2021). In the bigger picture, the ecological account of invariance also offers important insights into the nature of holistic harmony identification. By focusing on what is *constant* in harmonic experience even when other parameters (timbre, register, etc.) change, and by situating this constancy as an objective feature of the environment, the principle of invariance suggests a shift in pedagogical approach: The problem of developing holistic harmony identifications skills may not be so

much about learning to process and analyze unwieldy amounts of incoming sensory data, as it is about learning to detect and distinguish perceptual "meaning" objectively present in the stimulus. Once this has been recognized, pedagogical approaches explicitly founded in perceptual learning may be considered. This is the subject of the next chapter.

6 Rehearsal strategies for holistic harmony identification

6.1 Introduction

This chapter is an attempt to address Karpinski's main concern about "Gestalt listening"—renamed *holistic harmony identification* in this dissertation—namely that it is difficult to develop directly:

> Listeners at any given stage of development either recognize a particular chord or don't. In the absence of other listening strategies, they can be taught no concrete means for recognizing a "new" chord as a Gestalt. Instead, we might think of Gestalt listening as a by-product or result of other techniques. After weeks, months, or even years of repeatedly recognizing and labeling particular chords, those chords can become instantly recognizable (Karpinski, 2000, p. 119)

Based on the foregoing exploration of the term, I will use this chapter to propose and discuss some possible approaches to training holistic harmony identification. The categories presented here do not necessarily amount to the "concrete means" that Karpinski calls for, meaning they do not constitute specific training methods. Instead, they are broad rehearsal strategies that are meant to aid and inform more specific pedagogical approaches.

The rehearsal strategies are not intended as *alternatives* to Karpinski's approach of developing holistic harmony identification skills through repeated exposure. Extensive repetition is likely an inevitable element in such training: In an overview of the components of a cognitive-based approach to aural training, based on perceptual research, Herbst emphasizes that memory plays an important role in aural processing, and that repetition is a crucial part of forming memories (2021, p. 193). More efficient than simple repetition, however, is *elaborative* rehearsal, "where the material rehearsed has some meaning in relation to something already in long-term memory" (Snyder, 2000, p. 58). The strategies proposed here are meant to enhance a repetitive approach by introducing an element of such "meaning," thereby elevating it from possibly mindless repetition which may or may not eventually result in holistic harmony identification skills.

This is not to suggest that the strategies I propose are entirely novel. In the two decades since the publication of Karpinski's book, both research and aural training literature has emerged

that are relevant to the question of holistic harmony identification—even specific techniques, such as the *Do/Ti Test* (Stevens, 2016, 2020). My contribution in this chapter is therefore twofold: To make the explicit connection between holistic harmony identification and existing knowledge and teaching approaches, and to complement these with more novel suggestions.

6.2 The basic elements: Categorization and nomenclature

6.2.1 Categorical perception and learning

DeBellis suggests that aural training is "first and foremost the acquisition of a certain repertoire of perceptual concepts for musical events and properties, where one initially has no perceptual concepts of them" (2005, pp. 55–56). While I am going to nuance this view somewhat, I think it makes a good starting point for the present discussion: It seems reasonable to surmise that harmonic material cannot be identified holistically if one does not have a perceptual category for it. In fact, even basic perception is influenced by the categories possessed by an observer, an effect known as *categorical perception* (Goldstone & Hendrickson, 2010). Research on categorical perception has shown, for example, that the ability to discriminate between colors—both perceptually and in memory—is affected by one's native language, and whether or not that language contains distinct linguistic categories for the colors in question (Winawer et al., 2007). Therefore, verbal category labels are believed to play an important role in perceptual discrimination abilities, although it is not the only factor (Hendrickson et al., 2010). This effect is also believed to transfer to a musical context (Sloboda, 2000).

These insights have implications for aural training. Many traditional activities and learning goals in aural training are centered on the acquisition of perceptual categories for musical structures. Andrianopoulou proposes that a "strengthened categorical perception combined with richer and more conscious mental representations" may be a central part of aural expertise (2019, p. 89). An integrated part of such perceptual learning is the verbal *labeling* of the same structures. I have argued earlier in the dissertation (Chapter 4.1.1) that mastering theoretical labels is not a *necessary* condition for identification.[66] Still, it is likely that linking a language-specific component (i.e., a verbal label) to a musical phenomenon may assist in strengthening the perceptual category for that phenomenon—an idea supported by Winawer et al.'s study (2007). To illustrate this point, I will indulge in a personal anecdote.

66 I used the example of a rock guitarist who is able to play back the chords of a song after hearing it once, without being able to represent it symbolically (e.g., using chord symbols or Roman numerals).

While listening to Janáček's *Taras Bulba* (1915–1918) some years ago, I got hung up in a rather striking chord combination that appears multiple times throughout the last movement of the work. I clearly had no perceptual category for this specific chord combination, as I was not able to relate it to any type of chord combination I had heard before, in any genre of music. As a result, I was also unable to identify it holistically. What I *was* able to identify, was that the first of the two chords had the structure of a dominant 13 chord, and that the second chord was a tonic. What made the juxtaposition unusual was the fact that these two chords obviously did not express a dominant–tonic relation, i.e., the "dominant 13 chord" was not the V of the I chord, as is normally the case. This insight marked the boundaries for my holistic listening approach, and I had to start analyzing more consciously what I was hearing. I identified the descending fourth between the roots of the two chords, the $\hat{2}$–$\hat{1}$ contour of the melody, as well as the characteristic movement from $\flat\hat{3}$ in the first chord to $\natural\hat{3}$ in the second chord. Combining these pieces of information, I was able to conclude that the two chords must be as in Figure 8 (shown here in F major):

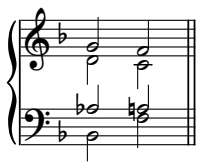


Figure 8: Moravian cadence

The point of this anecdote is that following this experience, I turned to publications on the harmony in *Taras Bulba*, and discovered that this chord combination has sometimes been termed a *Moravian cadence* (e.g., Zouhar, 2013). The chord combination now metamorphosed in my mind from the quirk of an individual composer into a clear harmonic concept. I had understood its structure, and it had, importantly, been coupled with a *name*. Following this incident, the Moravian cadence would perceptually "pop out" from its harmonic surroundings as I repeatedly encountered it in the music of Janáček's younger countryman Martinů, who uses it in a range of works (for example, in the last movement of his *Third Symphony* [1944]).[67]

My experience with the Moravian cadence is not unique, however: The aural identification and labeling of cadence types is a staple of traditional aural training. In a review of Karpinski's (2000) book, Rogers (2001, p. 90) argues that cadential recognition is in fact a form of beginner's "Gestalt listening" training. She advises starting by contrasting full and half cadences, then gradually introducing and refining recognition of predominants. Even Karpinski himself suggests that "listeners can be acclimated to hearing various cadence types … *making specific*

67 Hindle and Godsil (2006, p. 1090) suggest Martinů also coined the term "Moravian cadence."

Gestalt identifications with the terms 'authentic,' 'plagal,' 'deceptive,' 'Phrygian,' and even 'Landini'" (Karpinski, 2000, p. 140, emphasis added). Yet he seems surprisingly oblivious to his own argument when, earlier in the book, he laments that students "can be taught no concrete means" for learning to recognize new harmonic material as a Gestalt (Karpinski, 2000, p. 119). It is possible that Karpinski simply does not view cadential recognition as a "concrete means." I argue that it *can* be, insofar as holistic identification is made an explicit learning goal.

This point is obviously not limited to cadential archetypes. Learning to provide labels for perceptual objects is a common goal of aural training. This is my criticism of DeBellis's argument (referred to in the opening of this section) that aural training is "first and foremost the acquisition of a certain repertoire of perceptual concepts for musical events and properties, where one initially has no perceptual concepts of them" (2005, pp. 55–56). I will argue that this is actually only *part* of what aural training generally strives for. While the acquisition of perceptual concepts is essential where there are none in the first place, well-functioning (but perhaps unarticulated) perceptual concepts may be reinforced and expanded by fusing them with music-theoretical concepts (i.e., verbal labels), making the knowledge explicit: "Once tacit knowledge becomes explicit, arguably a loop is formed in which implicit knowledge would become automated knowledge once it reaches the explicit phase" (Herbst, 2021, p. 192). This process relates to Swanwick's (1994) model of musical knowledge. In this model—which builds on the work of Italian philosopher Benedetto Croce—intuitive and analytical knowledge are not dichotomies. Rather, intuitive knowledge is seen as a *prerequisite* for analytical knowledge and must always form the first step of the learning process. While DeBellis's idea of conceptual and non-conceptual listening can be interpreted in a similar manner, Swanwick goes further in explicating the hierarchical relationship between the two aspects of knowledge he presents:

> Propositional knowledge enters the dynamic process of musical knowing when it provides a vocabulary and a framework for secondary analysis. If the parameters of musical knowledge are adequately represented, then secondary analysis—the usual meaning of the term "analysis"—is an activity which can enlarge rather than constrain intuitive response. (Swanwick, 1994, p. 43)

In an aural training context, *intuitive knowledge* seems to translate broadly to perceptual familiarity (cf. perceptual concepts), while *analytical knowledge* concerns how this familiarity is categorized, systematized, and expressed verbally. Importantly, however, the development of analytical knowledge is not the end goal of learning in Swanwick's model. The two types of knowledge engage in a dialectic reciprocity, where analysis forms the basis for new and expanded intuitive knowledge, which again can be developed into analytical knowledge in a

continuous spiral motion. This reciprocity is the driving force of heightened musical awareness (Swanwick, 1994, p. 86ff).

6.2.2 Chunking

A unifying concept for the learning processes discussed above is possibly *chunking*. Snyder (2000) defines chunking as follows:

> A way of reducing short-term memory load by coding at a higher level. Chunks are small groups of elements (5–9) that, by being frequently associated with each other, form higher-level units, which themselves become elements in memory … Chunking is essential to the hierarchical organization of information, which makes its mental representation in memory much easier and more efficient. (p. 257)

The range 5–9 refers to Miller's (1956) classic article on "the magical number seven, plus or minus two," which he argues is the number of objects an average person can hold in short-term memory. Chunking allegedly bypasses this limit by recoding bits of information into larger units, or "chunks"—for example by memorizing a telephone number as four double-digits instead of eight single digits. The long-standing assumption that chunking reduces working memory load has gained empirical support, for example in a recent study by Thalmann et al. (2019). Though it is regarded as a mechanism of short-term and working memories,[68] chunking is also believed to connect to and interact with contents of long-term memory (Kurby & Zacks, 2008; Snyder, 2000; Thalmann et al., 2019), and is frequently characterized as an central component of musical memory (e.g., Johnson, 2013; Karpinski, 2000; Lehmann et al., 2007). Karpinski goes as far as considering it one of the *goals* of aural training, rather than merely a tool (2000, p. 77). I share this view, and I will even go further: I think it is difficult, if not impossible, to draw a clear line between well-developed chunking and holistic harmony identification. I have defined holistic harmony identification as the ability to detect, recognize, and differentiate relevant harmonic properties of a musical percept by means of its overall quality. Harmonic chunking does not *necessarily* adhere to this definition—it may be based on more concrete aspects of the musical percept than its "overall quality." Karpinski implies this when he claims that "listeners who chunk are thinking *analytically*" (2000, p. 77, emphasis added). Nonetheless, I can see no reason why harmonic chunking may not be *holistic*, that is, based on the recognition of emergent features of the harmony. On the contrary, I contend that chunks like "Phrygian cadence," "Moravian cadence," "descending fifth sequence," and so forth *are* emergent features. They are qualities that arise from the juxtaposition of certain

68 "Short-term memory" and "working memory" is understood here as the *maintenance* and the *maintenance plus manipulation* of information respectively, following a definition offered by Schaefer (2017, p. 33).

chord types. And here, in my opinion, harmonic *chunks* become indistinguishable from harmonic *Gestalts*, for all practical purposes. Therefore, while I might have misinterpreted what Karpinski means by "analytically" in the quotation above, I think he misses an important point regarding chunking: A well-developed ability to chunk harmony *relieves* the working memory of analytical effort—or, at least, redirects analysis from lower-level to higher-level structures. We could, thus, describe effective harmonic chunking using the same characterization that Karpinski offers for Gestalt listening: "Harmonic groups … drop out of the class of stimuli that require intellectual scrutiny in order to be perceived and take a place in a personal pantheon of essentially instantaneously recognizable entities" (2000, p. 119). This resemblance is also shown in the fact that Karpinski, in different parts of his book, holds up both chunking and "Gestalt listening" as end-goals of aural training. Still, as far as I can see, he makes no attempt to compare these two "techniques," or to distinguish clearly between them. Again, I think such a distinction might not be practically possible, which is why I suggested already in Chapter 1 that Karpinski's "Gestalt listening" (i.e., holistic harmony identification) may not be *one thing*, ontologically speaking. Therefore, it is possible that holistic harmony identification is best viewed as an umbrella term for ontologically varying forms of listening that nonetheless share a common outcome: that harmonic properties are detected, recognized, and differentiated by means of the *overall* quality of the percept. The relationship between harmonic chunking and holistic harmony identification should be addressed in future research.

6.2.3 Harmonic schemata

The discussion has so far focused on how traditional approaches in aural training may be tweaked to target the development of holistic harmony identification skills. There is another topic, however, that also merits a discussion here: harmonic schemata. In psychology, a *schema* is a representation of information in long-term memory (VanHandel, 2020, p. 5). *Event schemata* include "knowledge about what will happen in a given situation and often the order in which the individual events will take place" (Mandler, 2014, p. 14). *Harmonic schemata* are often understood as "stock musical phrases" used in a particular style (Gjerdingen, 2007, p. 6), and denote specific combinations of harmonic components[69] and the order in which they appear. Thus, the word "schema" in this context is used in a somewhat broader sense, pertaining to both objective musical structures and the aural identification capacities of a listener. While research on harmonic schemata often emphasize the former, there is also a growing awareness around how its findings may be applied in the aural training classroom (Peebles, 2021).

69 Often, "harmonic components" will be synonymous with "chords" for all practical purposes. However, some harmonic schemata—for example those presented by Gjerdingen (2007)—are more abstracted models of contrapuntal movement, and calling them "chord progressions" would be an oversimplification. Therefore, I use here the somewhat curious term "harmonic components."

There has been an increased musicological interest in harmonic schemata over the last decades. Gjerdingen's (2007) volume on schemata in the galant style of the eighteenth century marks a landmark and exemplar in this development, but treatises can also be found on the schemata of pop and rock music (Doll, 2017; see also Acevedo, 2020 and Hughes & Lavengood, 2021) and jazz (Shanahan et al., 2012).[70] Coker et al. (1997) should also be mentioned in this context. Although their book is not a "traditional" book on harmonic schemata—it is more concerned with the development of harmonic listening abilities than with systematically mapping harmonic practices—it presents and discusses common harmonic devices found in a corpus of approximately 500 jazz standards.

The harmonic schemata of some musical styles transfer to actual chord progressions. This is perhaps especially true of popular music, where popular schemata include "Doo-wop," "Singer/songwriter," "Puff," and "Double plagal" (Hughes & Lavengood, 2021). In European classical music, schemata are often more abstracted and idealized prototypes of melodic-harmonic figures, such as those presented by Gjerdingen (2007). This does not make them less relevant for aural training. Gjerdingen demonstrates this succinctly by laying out the harmonic structure of the second half of a slow movement by eighteenth-century composer Baldassare Galuppi:

> Quiescenza, diatonic, repeated
> Fonte/Monte combination
> Ponte, to Passo Indietro
> Comma, followed by Cudworth cadence
> Clausula Vera
> Meyer
> Ponte, tonic
> Monte/Converging cadence combination
> Fonte, repeated
> Monte, diatonic
> Clausula Vera
> Ponte
> Cudworth cadence … deceptive
> Passo Indietro to Mi-Re-Do cadence
> (2007, pp. 7–8)

These labels, naturally, mean nothing to a person unfamiliar with them. Yet they all capture and conceptualize schemata frequently found in the galant style—and beyond: Schemata like the

70 In a more traditional music theory textbook, Laitz (2016, p. 246) uses the term "harmonic paradigms" in much the same way as other scholars use the term "harmonic schemata."

"Romanesca" and the "Prinner" have been staples of tonal harmony since they first appeared, although they have not necessarily been known by these names. Some of the labels stem from contemporary use in the Neapolitan pedagogical tradition, some have been introduced by later theorists, and some are coined by Gjerdingen. The value of collecting and presenting them in a systematic manner is not only that it expands our theoretical insight into the compositional processes of music in the galant style, but that it enables the development of "packet[s] of knowledge" (Gjerdingen, 2007, p. 11) in an avid listener. It offers a verbal repertoire for harmonic phenomena that might otherwise remain abstract or unarticulated. This repertoire may then be practiced aurally the same way that cadential recognition is traditionally practiced in aural training—to the point where schemata can be recognized holistically. Peebles (2021, p. 320) argues that Gjerdingen's galant schemata can be interpreted as a chunking strategy on a larger scale. This again demonstrates the blurry line between well-developed harmonic chunking abilities and holistic harmony identification.

Finally, I will include a book from the jazz tradition. In *Hearin' the Changes: Dealing with Unknown Tunes By Ear*, Coker et al. (1997) seek to develop the readers aural familiarity with common harmonic traits in jazz music. The authors promote what they call "associative listening," which consists in noticing, observing, and mentally "tagging" the sound of such chord combinations. While presented as a "technique" (pp. 8–9), it is perhaps more correct to describe associative listening as a general attentiveness to the perceptual quality of chords and progressions, and to how they "make us feel" (p. 35). The goal is to facilitate *identification by association*, which connects sensory information to patterns in the memory bank. Thus, Coker et al. synthesize many of the concepts treated in the foregoing discussion—categorization, labeling, chunking, and harmonic schemata—all presented in an aural training context.

To summarize, I believe that the growing focus on harmonic schemata in music theory research may offer ways to expand aural training. It also offers new repertoires of harmonic concepts that may eventually "take a place in a personal pantheon of essentially instantaneously recognizable entities" (Karpinski, 2000, p. 119).

6.2.4 Discussion

In conclusion, "concrete means" for practicing holistic harmony identification may not be so unattainable as Karpinski portrays them. Well-established aural training activities can be calibrated towards this goal. For example, the customary recognition and labeling of harmonic material (e.g., "Phrygian cadence" or "descending fifth sequence") that is practiced in most aural training classrooms is arguably already a form of "novice" holistic identification training. It may facilitate and strengthen the perception of those harmonic formulae as "chunks,"

thereby enabling the development of holistic identification abilities. The growing research interest in harmonic schemata represents a potential expansion of the repertoire of stock harmonic patterns to be familiarized in aural training.

Hence, the pedagogical challenge is not necessarily to invent novel exercises completely, but to introduce an explicitly holistic listening *approach* into existing exercises. This is also pointed out by Jarvis (2015) in an essay on holistic harmony perception. Jarvis argues that aural training students are in fact already hearing a lot of harmony holistically, and that teachers can take advantage of and nurture this ability through a conscious teaching approach. Though he advocates statistical learning through massive repetition of the harmonical material, Jarvis is also attentive to letting students experience and contemplate the specific sound quality of different chords. In the next section I will discuss how such listening may be strengthened.

6.3 The glue: Metaphorical listening

6.3.1 Introduction

Snyder suggests that *associations* are the glue that holds the elements of a chunk together (2000, p. 54). Drawing on research on categorical perception, I have already suggested that verbal labels may strengthen a listener's ability to perceive chord combinations as *one* perceptual object, or chunk. This is the learning goal of traditional aural training activities like cadential recognition. The theoretical labels used in such activities—whether "Phrygian cadence," "Romanesca," or "back door progression"—perform the function of the "glue" that Snyder describes: a unifying concept of which individual chords become a part; a prototype for a harmonic device that one can learn to recognize aurally. In this section, I will explore the concept of association more broadly, not limited to music-theoretical labels.

6.3.2 The nature of labels

First, let us consider whether verbal labels must necessarily be *theoretical* to fulfill the purpose described above—that is, to function as a linguistic "hook" for a perceptual phenomenon. I argue that they do not, and for two reasons. First, I have already discussed the "James Bond chord" (i.e., a minor major ninth chord). This is clearly a non-theoretical label, yet it efficiently captures a salient perceptual quality of that chord, namely its cultural association with spy movies and the James Bond theme in particular. Another example is the "Hendrix chord" (i.e., an augmented ninth chord)—or, as my former guitar teacher Jan Fredriksen dubbed

it as a young teenager with limited theoretical understanding: the "pig seven" chord.[71] The argument for music-theoretical labels, naturally, is that they are part of a larger discourse. The term "augmented ninth chord" conveys structural properties of the chord, as well as being part of a shared, standardized nomenclature in an international music community. For the purpose of aural recognition and consolidating the chord as a concept in the consciousness of an individual listener, however, the labels "Hendrix chord" or "pig seven" are no less apt. For similar reasons, Tagg advocates the use of a more inclusive nomenclature in music education:

> I've found that many students, muso and non-muso, can, if their attention is drawn to the sonority in question, recognise not only detective chords but also other aesthesically labelled sonorities like the BITTER-SWEET CHORD, the ROMANTIC PATHOS CHORD and BURT BACHARACH CHORDS. They can also usually distinguish between drone-based and busily over-harmonised arrangements of folk tunes, between the hormonic idioms of trad jazz and bebop, between Elizabethan and late Romantic harmonies, etc. The problem is in other words not one of aural competence but of poïetic nomenclature (like 'minor major nine' ($m^{\Delta 9}$)) because aesthesic descriptors like BITTER-SWEET, ROMANTIC PATHOS, BURT BACHARACH and TWANGY FOLK chords have (as yet, if ever) little or no validity in institutions of conventional musical learning. (Tagg, 2013, p. 340)

The second reason labels need not necessarily be theoretical is that "theoretical" is a vague category. For example, when introducing the term "Cowboy cadence" to denote a subtonic-to-dominant half cadence (♭VII–V), Tagg and Clarida (2003) lean on the same kind of popular association as Tagg did in the quote above. The cadence was important in establishing the sound of Hollywood Western movies in the 1950s—hence the colloquial term "Cowboy cadence." The adoption and elaboration of the term by other musicologists (e.g., Lehman, 2013, 2018; Schneller, 2013), however, has arguably made it transition into an established theoretical label. Therefore, the distinction between "theoretical" and "non-theoretical" labels may not always be detectable. Nor is it important, in my opinion. In the context discussed here, labels have a practical purpose—functioning as a verbal "anchor" for elusive perceptual information.

6.3.3 A broader metaphorical approach to listening

I have already touched upon several concepts that could be considered "metaphorical" approaches to harmonic listening, including a discussion of harmony's extra-musical connotations (see 5.3). These may include for example cultural associations (e.g., the "Wedding chord"), associations to mass media tropes or musical topics (e.g., the "spy chord" or "Cowboy cadence"),

71 The term's delicious ambiguity and pun-like quality in Norwegian, *gris-sju*, is difficult to convey in English.

or associations to specific musical contexts (e.g., "Hendrix chord" or "Let It Be-progression"). They may also include associations to physical (e.g., "sharp" or "soft," "warm" or "cold") or affective qualities (e.g., "romantic," "bitter-sweet," "wistful," or "tense"). Yet others include auditory-motor associations (e.g., "hearing" how it feels to play a chord on the piano, or where it is placed on the guitar neck) or visualization (e.g., how a chord looks on a keyboard, or visualizing chord symbols). In addition to these, numerous highly personal and perhaps also non-linguistic associations are conceivable.[72] Within an ecological framework, all of these associations may be seen as perceptual meanings afforded by the stimulus. That is, they are all results of a listener trying to make sense of the perceptual information available in the sonic environment.

Rather than suggesting a hierarchy, I propose that *all* such associations can be put to good use in the development of holistic harmony identification skills (although not all are equally easily implemented in the aural training classroom). In a study on the identification of well-known jazz standards from isolated chord progressions, Jimenez and Kuusi (2018b) found that identification was greatly facilitated when the participants had both multi-sensory (auditory, motor, and visual) and multi-domain (perceptual and conceptual) experience with the chord progression in question. This supports the idea of using a broad approach to practicing harmonic recognition, incorporating as many senses, activities, and forms of understanding as possible. I call this a "metaphorical" listening approach for two reasons. First, to distinguish it from the somewhat narrower "associative listening" described by Coker et al. (1997) and Jimenez and Kuusi (2018b). Second, because I find that conceptual metaphor theory and the idea of cross-domain mapping (see Part II of this dissertation) offer informative theoretical frameworks for further explorations. In these research traditions, the *metaphor* is a central idea. I am not the first to suggest the relevance of metaphor to aural training. Andrianopoulou stresses the importance of metaphor in exploring musical meaning and expressing it verbally, as well as serving as "the link that bridges the structural features of a musical work with its wider context(s), connecting intra-musical with extra-musical perspectives of musical meaning" (2019, p. 101).

If a "metaphorical" listening approach seems vague, it is because it does not involve a specific curriculum, or specific techniques. Rather, it is about fostering a rich, holistic, and phenomenologically oriented way of listening which is attentive to all the potential meanings that sounds may afford. This, I propose, is a better strategy for developing holistic harmony identification skills in students than the atomistic, interval-oriented, and emotionally detached methods that can often be found in traditional harmonic aural training. A related sentiment

72 While the examples listed here are hypothetical, they reflect the findings of my master's thesis on professional jazz musicians' harmonic listening (Langfeldt, 2017).

is expressed by Andrianopoulou, who contends that the competencies and techniques that aural training seeks to develop cannot be seen in isolation from "memory, implicit knowing, embodied knowing, social interaction, and emotions" (2019, p. 172). Aural training recast as aural *education* must

> assimilate as far as possible the full richness of the musical experience … helping students not only to develop practical musical skills, but also to become more aware of what they already know musically, and of the mechanisms through which they know it. (Andrianopoulou, 2019, pp. 172–174)

More research and experiences from practical aural training are needed to develop the idea of a "metaphorical" approach further. As a first contribution, I will use the last two parts of this dissertation to explore various notions of "metaphor" in harmonic listening.

7 Summary

In this theoretical study, I have examined the concept of "Gestalt listening," introduced by Gary S. Karpinski (2000). While promoted as an expert skill and an end-goal of harmonic aural training, the concept has suffered from the lack of a clear definition. As a result of this, it has not been clear whether it can be trained intentionally, or if it only evolves inadvertently from repeated, long-term exposure to harmonic material.

I have suggested that Karpinski's choice of term, "Gestalt listening," is problematic. Gestalt theory is obviously relevant to the concept, but Gestalt listening is not about the basic perceptual grouping of simultaneous pitches into auditory objects (i.e., chords), which is the most straightforward understanding of "harmonic Gestalts" (cf. Bregman, 1990, p. 496). Rather, Karpinski is concerned with higher-lever Gestalts, and how these are holistically *identified* by a listener. Furthermore, the word "listening" arguably implies a certain passiveness on the listener's behalf.

For these reasons, and in order to clarify and better define the concept (cf. research question 1), I have proposed the term *holistic harmony identification*—a term used elsewhere by Karpinski himself. I have defined it as follows:

Holistic harmony identification is to detect, recognize, and differentiate relevant harmonic properties of a musical percept by means of its overall quality—i.e., its emergent features—and to be able to express these properties through a corresponding, external response. Harmonic properties in this regard are pitch simultaneities or successions which afford perceptual grouping into chords or chord progressions.

Combining this definition with an ecological or "Gibsonian" framework, I have furthermore examined the perceptual basis for holistic harmony identification, how it might be experienced from a listener's perspective, and its role in the aural training classroom (cf. research question 2). This examination led to the hypothesis that a "metaphorical" listening—an awareness of the implicational meanings that harmony may afford—possibly offers a way of practicing holistic recognition of harmony. Combining this listening approach with existing practices in aural training, I have finally suggested some broad strategies for such targeted training (cf. research question 3).

The remaining parts of the dissertation will focus on different aspects of a "metaphorical" listening approach.

PART II

Harmonic metaphor mining: An explorative, qualitative document analysis of cross-domain mapping in literature on musical harmony

8 Introduction

8.1 Background

In Part I of this thesis, I explored the concept of holistic harmony identification, portrayed by Karpinski (2000, p. 425) as an expert harmonic listening skill. In a discussion of possible rehearsal strategies for developing this skill, I proposed a "metaphorical" listening approach. Broadly speaking, this involves perceptual awareness of harmony's implicational meanings, for example its ability to specify other domains than that of music. I discussed two rather salient examples of single chords that convey "meaning" in this way: The "Mendelssohn chord" and the "James Bond chord," which, at least within a Western cultural sphere, carry rather strong connotations to "wedding" and "spy" respectively. Nonetheless, a handful of examples like these do not by themselves make a convincing argument for the metaphorical listening approach I have proposed. Further explorations of the role of metaphor in harmonic perception are necessary.

There is a long and still thriving tradition for exploring various notions of metaphor in music, spanning vastly different approaches (e.g., Agawu, 1991; Brower, 2000; Cooke, 1959; Davies, 1994; Eitan, 2017; Hatten, 1995; Johnson & Larson, 2003; Larson, 2012; Meyer, 1956; Saslaw, 1996; Tagg, 2013; Walther-Hansen, 2020; Zbikowski, 2017a). However, I concur with Cooke's sentiment that the meanings of music cannot be formulated in a dictionary (Cooke, 1959, p. 13). I do not believe that it is possible to create some kind of complete repertory of individual chords and chord progressions and their metaphorical quality—at least not one that most individual listeners would agree on. Therefore, no such attempt will be made in this dissertation. Instead, the present study is an examination of existing metaphorical mappings in Western conceptualizations of musical harmony and its perceptual effect.

8.2 Metaphors we hear by

Whenever we want to discuss, compare, or share our perceptual experiences with others, we are left with few other options than trying to convey these experiences through the use of words. When we do this, we transfer something from one mode of experience (sensory) to another (linguistic). This transfer can highlight important facets of the experience, but it may also blur or distort other aspects of it. Either way, chances are the verbalization will be

a *metaphorical* representation. Try describing the color yellow without using metaphorical expressions like "energetic," "sunny," "bright," "warm," "tangy," "lemony," and so forth.

Imperfect and second-rate reproductions of the actual phenomenon though they are, such verbalizations are not necessarily arbitrary. This can be observed in activities requiring acute perceptual discrimination. Often, such activities develop a highly detailed vocabulary to capture perceptual subtleties. Yet this vocabulary may still correspond closely to objective, material properties of the stimulus. For example, wine tasters may refer to aromas of freshly cut grass and passion fruit in a Sauvignon Blanc. These are obviously metaphors—the wine does not actually contain grass or passion fruit—but as descriptors, they also correspond to the chemical compounds *hexanal* and *3-mercaptohexan-1-ol*. These compounds are commonly present in wine made from Sauvignon Blanc grapes—and in freshly cut grass and passion fruit respectively (Krebiehl, 2020). In a completely different area, Ryd (2022) documents over 300 Lule Sámi terms concerning snow and ice. The terms concretize phenomena, conditions, and situations that might otherwise be left undetected, for instance *oarrevahtsa*, "squirrel-track snow": A minute layer of recently fallen snow which enables the hunter to distinguish the freshest tracks from the slightly older ones (Ryd, 2022, p. 145). Ryd's example illustrates Hatfield's (2014) point that an organism's activity in relation to the surroundings *generates* stimulus information (p. 365). In the area of music perception, Walther-Hansen (2020) has showed that common metaphors for musical sound quality—e.g., *balanced*, *warm*, *bright*, *wet*, *smooth*, *tight*, *fat*—generally correlate to objective physical features of the sound. This is the central claim of his book: Common metaphors in descriptions of sound arise because they closely resemble the auditory experience.[73]

The language we use to describe musical harmony is also rife with metaphors. Indeed, the very concept of "harmony" is what Lakoff and Johnson (1980) call an *ontological metaphor*: viewing something as an entity, so that we can more easily refer to it, quantify it, identify aspects of it—in short, better understand it. A skeptic might raise the argument that musical harmony is nothing but a range of psychoacoustic effects, caused by sound stimuli of different wavelengths reaching our auditory system simultaneously. However, the Western music tradition has long treated it as something that exists in and of itself. Harmony has rules and principles; it has grammar and syntax; it can be used consciously to convey a range of emotional states in a range of musical styles; it has cultural and semiological meanings that can refer to something outside itself; it can be examined and analyzed; and, not least, it can be studied and taught, as in *the study of harmony*.

73 A similar sentiment has been expressed by philosopher Roger Scruton (1983), although Scruton goes even further: "If we take away the metaphors of movement, of space, of chords as objects, of melodies as advancing and retreating, as moving up and down—if we take those metaphors away, nothing of music remains, but only sound" (pp. 84–85).

But harmony *is* and *does* much more. And in describing what it is, what it does, and how we experience and understand it, we employ a range of metaphors. Some of these are so integrated in modern music terminology that they have become "dead" metaphors—we hardly notice their metaphorical qualities. Container metaphors are used to define the contents of scales, chords, keys, forms, etc. Structural metaphors—structuring harmony in terms borrowed from a different concept—allow us to talk naturally about *weak* harmonic beats or functions, or of *strong* cadences. Some metaphorical descriptors are, perhaps, more noticeable—like when we appreciate *smooth* voice leading, criticize an *abrupt* or *jarring* modulation, analyze a *deceptive* chord resolution, or describe the *struggle* or *conflict* between two keys in sonata form. Dead metaphors like these are relatively uncontroversial, and their meanings are both accessible and grounded in objective perceptual experience. On the other hand, there is no doubt that the perception of musical harmony also includes a subjective dimension. When French composer Olivier Messiaen describes harmonies as *swords of fire*, *sudden stars*, *flows of blue-orange lavas*, *planets of turquoise*, or *garnets of long-haired arborescence* (Messiaen, 1990, p. 52), we must trust that he does indeed experience harmony this way.[74] However, it is not likely that very many others do. Likewise, when his Czech colleague Leoš Janáček proposes that the 6/4 chord inversion sounds like "the swallow flying which almost touches the ground" (cited from Beckerman, 1994, p. vii), it is doubtful whether this conceptualization offers any help to the struggling aural training student. What is of interest to the context of aural training, in my opinion, are the more recurrent metaphors found between the opposite poles of dead metaphors and complete idiosyncrasy. Adapting Walther-Hansen's claim, I postulate that such metaphors, too, generally resemble the auditory experiences of which they are verbalizations.

This study examines metaphorical descriptions of harmony, by analyzing how they are used in a variety of texts, by a variety of authors. The goal is to get a deeper understanding of how metaphors are used to structure and verbalize the subjective experience of perceiving harmony. As proposed in Part I of this dissertation, metaphorical listening strategies might offer a pathway to practicing holistic harmony identification for students who do not possess such aural skills. Attunement to metaphors, and to the perceptual qualities they describe, is a possible way of enriching one's perceptual awareness during harmonic listening. The pedagogical relevance of the metaphors identified in this study will be discussed.

74 And, considering Messiaen's famous synesthesia (Austbø, 2015), it is probably not unlikely.

8.3 Earlier research on music and metaphor

As already mentioned, there is a long and varied research history on different notions of metaphor in music. I will limit myself to mentioning here those that bear particular semblance or relevance to the present study.

First, several studies have drawn on different aspects of conceptual metaphor theory to examine music perception. Conceptual metaphor theory (see Chapter 9.1 for a more thorough introduction) originally emerged from cognitive linguistics and is strongly associated with the work of George Lakoff and Mark Johnson. Two main tenets associated with conceptual metaphor theory can be mentioned in this context: First, that human thought is fundamentally metaphorical in that we structure abstract concepts by "borrowing" the structure of more familiar concepts (e.g., Lakoff & Johnson, 1980). This is known as *cross-domain mapping*. Second, that central to such mapping are *image schemas*—pre-linguistic patterns abstracted from, for example, bodily experience (Johnson, 1987).

Saslaw (1996) has showed that already Hugo Riemann relied heavily on metaphors to conceptualize of his function theory. These metaphors are visible in ideas central to Riemannian theory, e.g., CADENTIAL PROGRESSIONS ARE CONFLICT AND RESOLUTION, and lend structure to phenomena that are not easily expressed in auditive terms.[75] In a theoretical study, Brower (2000) has proposed that image schemas—such as CONTAINER, CYCLE, VERTICALITY, BALANCE, CENTER–PERIPHERY, and SOURCE–PATH—play an important role in the understanding of music. She has furthermore examined how they inform the mental organization of harmonic relationships and progressions. Gur (2008) analyzed the role of image schemas in the eighteenth-century harmonic theory of Rameau (1971) and found that two schemas in particular are essential: The PATH schema conceptualizes harmonic progressions as a goal-oriented and continuous process, whereas the FORCE schema structures dissonance and harmonic causality. Larson (2012) has showed empirically that metaphorical concepts like MUSICAL GRAVITY, MUSICAL MAGNETISM, etc. could be detected by listeners. The work most closely related to the present study, however—both in terms of theoretical framework and method—is that of Walther-Hansen (2020).[76] As already mentioned, Walther-Hansen examines the use of metaphors in the description of sound quality. Like the present study, he combines a conceptual metaphor theory framework with an analysis of written texts (from music reviews, hi-fi magazines, and sound engineering literature). The primary methodological difference between Walther-Hansen's study and the present one relates to corpus size:

75 Throughout Part II, conceptual metaphors and conceptual domains will be written in SMALL CAPS, following the tradition of conceptual metaphor theory.

76 The analysis for the present study was also performed in 2020, so I was not aware of Walther-Hansen's book at the time. I am thankful to Lawrence Zbikowski for bringing it to my attention during my trial disputation in 2021.

His corpus consisted of more than 50 million words. Therefore, the analysis was based on the use of software, which seems to necessitate certain hypotheses about which metaphorical expressions to search for. The present study is more exploratory and poses no hypotheses about which metaphors are likely to be found. All metaphors are recorded and analyzed.[77]

In a slightly different direction, the last couple of decades have seen an increase in empirical, psychological experiments on cross-domain mapping in music (e.g., Eitan & Granot, 2006; Eitan et al., 2012; Spence, 2019). This tradition has arguably branched off and formed an independent research paradigm, that of *cross-modal correspondences*. Research on cross-modal correspondences is more explicitly based in (subconscious) perceptual processing and the blending of sensory modalities. In the present study I will apply an inclusive approach to the term *cross-domain mapping*, one that also encompasses cross-modal correspondences (see discussion in the *Theory* section). Part III of this dissertation, however, is an empirical study on cross-modal correspondences, and will contain a thorough introduction to the research history and methods associated with this tradition.

8.4 Aims of this study

The main goal of this study is to further the awareness and understanding of metaphors commonly applied to the description of harmony. This will be done through a document analysis of 20 books about harmony. The analysis will lead to a presentation of common metaphor categories, shedding light on how harmonic phenomena are conceptualized through metaphorical descriptions.

The metaphors investigated in this study pertain to alleged esthesic effects of harmonic phenomena. It is *not* an aim of this particular study to assess the perceptual veracity or universality of these metaphors—only to record, analyze, categorize, and present them. However, one of the conceptual metaphors[78] analyzed here will be used for perceptual testing in a statistical study in Part III of this dissertation.

The pedagogical relevance of the study is to explore notions about *how harmony sounds* that could potentially be applied to aural training. A hypothesis presented in Part I of the dissertation claims that metaphorical descriptions of harmony are generally holistic descriptions—they describe something about the *overall character* of a chord, progression, modulation, etc.

77 See 10.4 for exceptions.

78 The term *conceptual metaphor* (cf. Lakoff & Johnson, 1980) will be introduced and discussed in Chapter 9.1.

Hearing a harmonic phenomenon "as" something necessitates focusing one's aural attention towards it, judging it as a whole, noticing its emergent qualities, and describing—in non-theoretical terms—something about its Gestalt quality. I have suggested that this approach encourages holistic, cross-modal and meaning-oriented ways of listening—in contrast to the atomistic, intervallic-oriented, and emotionally detached approaches of a more traditional aural training pedagogy. This, again, may enrich the perceptual experience and promote the aural skills associated with holistic harmony identification.

Accepting this hypothesis, metaphors harvested from the literature on harmony might represent pathways to an intentional training of holistic harmony identification skills. A lack of such training methods was highlighted by Karpinski (2000) as the primary weakness of this expert skill. As a banal example, consider the common emotional connotations of the major and minor modes, i.e., the notion that major expresses a "happy" quality, while minor is "sad". Keeping criticism of this simplified dualism aside, these descriptions conceptualize the difference in esthesic effect generally attributed to the two modes respectively. Using affective associations of "happy" and "sad," it is feasible even for a person without musical training to learn to distinguish rudimentarily—and holistically—between music in major and minor. Tagg (2013) describes a similar approach, whereby non-musicians are able to aurally connect relatively complex chords to esthesic descriptors like the James Bond chord, the bittersweet chord, or the romantic pathos chord (pp. 339–340).

A secondary aim of this study is to search for common metaphors across highly different sources. While superficial[79] associations like Tagg's James Bond chord are not excluded from this search, more significance is given to systematic mappings between harmony and another domain, i.e., conceptual metaphors. If the same conceptual metaphor structure can be identified in sources from different fields and traditions, it could indicate that that specific way of conceptualizing harmony is perceptually available to a larger population. Furthermore, even if a particular metaphor is found to be common in literature on musical harmony, this does not mean that it is being utilized in a pedagogical context. As the work of Lakoff and Johnson (1980) demonstrates, we are often oblivious even to conceptual metaphors that we use in everyday language.

Therefore, this study only represents a first step towards a metaphorical approach to aural harmony training. This first step is *awareness*. By laying harmonic metaphors out on the table and sorting them into categories, it will be easier to examine them and determine which ones may have a potential for helping students improve their harmonic listening skills. Then comes the work of exploring practical teaching methods that encourage holistic listening.

79 That is, more or less accidental associations; not expressive of a larger, underlying metaphor structure.

9 Theory

For the identification and analysis of harmonic "metaphors" in this study, I will lean on theory from two related—but not identical—research traditions. These are *conceptual metaphor theory* and research on *cross-modal correspondences*. The former is the main theoretical framework for the study, while the latter is primarily used as a supplementary perspective on one particular metaphor type.

Because they represent different traditions and approaches, I will present the two strands of research separately, before arguing for a combinatory approach.

9.1 Conceptual metaphor theory

Throughout this study, the words "metaphor" and "metaphorical" are used to denote a *cross-domain mapping* between music and another domain. Zbikowski (2002, p. 13) explains cross-domain mapping as "a process through which we structure our understanding of one domain (which is typically unfamiliar or abstract) in terms of another (which is most often familiar and concrete)."[80] The concept is closely associated with conceptual metaphor theory, and with the work of cognitive linguist George Lakoff and philosopher Mark Johnson.

Lakoff and Johnson (1980) maintain that metaphors are not just figures of speech, used to "spice up" our language. Rather, they are models of thought that shape our understanding of the world. In a classic example, Lakoff and Johnson demonstrate how the English language uses the concrete domain of MONEY to structure the abstract domain of TIME. This gives rise to the conceptual metaphor TIME IS MONEY. Conceptual metaphors are seldom expressed directly[81] but are implied through the use of more superficial metaphorical expressions. For example, TIME IS MONEY engenders a whole range of metaphorical expressions describing TIME with concepts borrowed from the domain of MONEY:

> You're *wasting* my time.
> How do you *spend* your time these days?
> That flat tire *cost* me an hour.
> Is that *worth* your while?

80 Cross-domain mapping is sometimes also referred to as *conceptual mapping*.

81 Although this is obviously not true in the case of TIME IS MONEY, which is so strongly associated with Benjamin Franklin and has in itself become something of a motto of the modern, capitalist world.

He's living on *borrowed* time.
(All examples from Lakoff & Johnson, 1980, pp. 7–8)

The point, in conceptual metaphor theory, is that these everyday expressions are generated from, and point back to, the conceptual metaphor TIME IS MONEY. This metaphor expresses a more fundamental idea that—particularly in a capitalist society—can permeate and shape our conception of what time *is*, and how we relate to it. Time, in this view, may be perceived as a scarce resource, and squandering it through laziness is therefore akin to wasting physical resources. However, we can easily imagine cultures where time is traditionally not seen as a limited asset—an hypothetical society living by the metaphor TIME IS LOVE would likely make very different priorities—illustrating that TIME IS MONEY is not necessarily a *universal* conceptual metaphor. This can be explained by the use, in this particular conceptual metaphor, of a cultural source domain (MONEY). However, many conceptual metaphors use sensorimotor source domains which stem from our embodied experiences of moving through and functioning in the physical world. Examples of such metaphors are DIFFICULTIES ARE BURDENS,[82] AFFECTION IS WARMTH,[83] IMPORTANT IS BIG,[84] and MORE IS UP[85] (Gibbs, 2017, pp. 29–30). Many conceptual metaphors based on recurring bodily experiences have been found to be largely universal (Gibbs, 2017, p. 39).

In spite of certain claims to universality, Gibbs (2017) maintains that pinpointing the role of conceptual metaphors in human understanding is not a straightforward task. Indeed, they "may be emergent products of multiple, nested factors (i.e., biological, historical, cultural, social, cognitive, and linguistic), and may interact with many knowledge sources and experiences to create context-sensitive, task-specific metaphorical behaviors" (p. 15). I will adopt this inclusive view in the present study.

The primary function of conceptual metaphor theory in this study is its explanatory model: Conceptual metaphors arise as conceptualizations of abstract phenomena or ideas. They are generally formed by mapping properties of a familiar source onto an unfamiliar target domain, thereby giving it a metaphorical structure. As a product of this mapping, a conceptual domain is created, in which for example "time" has properties normally associated with "money." What I expect to find in the data are metaphorical *expressions* (e.g., "that flat tire *cost* me an hour"). It will then be part of the analysis to examine whether these are expressions of underlying conceptual metaphors (i.e., TIME IS MONEY).

82 "He is carrying the *weight* of his father's sins," my example.

83 "She gave me a *warm* welcome," or "his attitude towards her is *cold as ice*," my examples.

84 "*Tremendous* news!," my example.

85 "Food prices are *rising*," my example.

9.2 Cross-modal correspondences

Cross-modal correspondences[86] describe "the widespread tendency for attributes in one sensory modality to be consistently matched to those in another modality. For example, high pitched sounds tend to be matched to spiky shapes, small sizes, and high elevations" (Hamilton-Fletcher et al., 2018, p. 114). Research on cross-modal correspondences is associated with perceptual psychology, and has different perspectives and research focuses than conceptual metaphor theory. The two traditions nonetheless share a common interest, namely the role of cross-domain mapping in human experience. A main difference, simply put, is that conceptual metaphor theory is concerned with how cross-domain mapping impacts thought, while cross-modal correspondence research is concerned with how it impacts perception.

A well-known example of a cross-modal correspondence is the *bouba–kiki effect* (figure 9). It exists in a number of variations, and was first demonstrated in an experiment by Gestalt psychologist Wolfgang Köhler (1929) almost a hundred years ago. In the most famous version of the test, the participant is asked to decide which of the figures is *Kiki*, and which is *Bouba*.

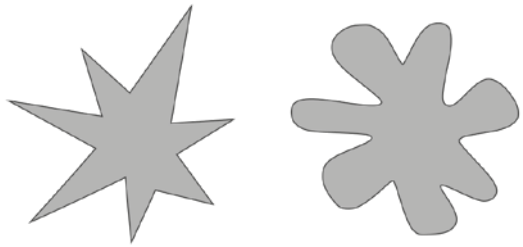


Figure 9: Kiki & Bouba (Wikimedia Commons)

The bouba–kiki effect consists in the significant general agreement that the figure on the left is Kiki, and that the one on the right is Bouba (e.g., Ramachandran & Hubbard, 2001). While different explanations for the effect have been proposed, e.g., that it has a semantic basis (Milan et al., 2013) or that it is influenced by orthography (Cuskley et al., 2017), most theories assume a cross-modal correspondence between the geometrical shapes (i.e., the visual domain) and the sound of the words (i.e., the auditory domain): the "sharp" shape is associated with the "sharp" word *Kiki*, while the "soft" shape is associated with the "soft" word *Bouba*. This explanation is supported by the demonstration of the bouba–kiki effect across

86 Sometimes referred to as cross-modal *integration*, cross-modal *association*, cross-modal *perception*, or cross-modal *analogy*. Cross-modal correspondences are not, however, the same as *synesthesia*, which is the involuntary triggering of an unrelated perceptual mode (see discussion in Part III).

ages (Maurer et al., 2006) cultures (Bremner et al., 2013) and writing systems (Ćwiek et al., 2022). An underlying assumption in much cross-modal correspondence research is precisely that such correspondences are generally shared by a large number of people, and that some may even be universal (Spence, 2011, p. 973).[87]

The primary function of cross-modal correspondences in the present study is to include an alternative view of cross-domain mapping to that of conceptual metaphor theory. In other words, the analysis will consider whether certain cross-domain mappings expressed linguistically in the source material are better explained as indicative of sensory modality integration, rather than of conceptual metaphors. However, in this part of the dissertation I will not assess the perceptual *veracity* of such mappings—i.e., whether or not they represent "true" cross-modal correspondences shared by a large number of people[88]—only to analyze which aspects of musical harmony are being highlighted through the mapping.

9.3 Cross-domain mappings: A combinatory approach

In this study, I will use the term *cross-domain mapping* to include mappings both across sensory modalities (i.e., cross-modal correspondences) and across cognitive/conceptual domains (i.e., conceptual metaphors). Although the term has traditionally been associated with conceptual metaphor theory, it has also been used in research on cross-modal correspondences—especially in the pioneering work of Zohar Eitan (Eitan, 2013, 2017; Eitan et al., 2012; Eitan & Timmers, 2010; Eitan et al., 2017). For example, Eitan et al. (2012) use the term *cross-domain mappings of sound* to refer to both cross-modal correspondences and "pitch metaphors" like *heavy*, *old*, *thick*, *happy*, etc. These metaphors are treated as descriptions of physical properties like size, energy, and sharpness, and conceptualized within a cross-modal correspondence framework. However, it appears to me it would also be possible to explain them using conceptual metaphor theory, which also leans heavily on image schemas derived from physical experience.

Furthermore, research on cross-modal correspondences is arguably in a state of development, with tentative and fluid understandings of what cross-modal correspondences *are* (see Part

87 To avoid giving the impression that the bouba–kiki effect is universal, I should note that it has been found to be weaker in at least four populations: in a population in Papua New Guinea (Rogers & Ross, 1975); in people on the autism spectrum (Oberman & Ramachandran, 2008; Gold & Segal, 2017); in blind and partially sighted participants (using haptic reproductions of the shapes; Fryer et al., 2014); and in participants with prelingual hearing loss whose hearing had later been rehabilitated (Gold & Segal, 2020).

88 This is the aim of Part III of the thesis.

III of this thesis for a more detailed discussion). This is reflected in Spence's suggestion that they may involve both "low-level amodal stimulus properties such as duration [and] high-level cognitive correspondences based on stimulus meaning/valence" (2011, p. 973), and in a question raised by Parise:

> A fundamental question in the field is whether the different phenomena labeled as crossmodal correspondences are indeed a unique phenomenon, or whether such compatibility effects reflect a gamut of different underlying associative processes. While some correspondences, such as the simple associations across basic features from different senses, might rely on perceptual mappings, other correspondences involving more complex stimuli (such as words and complex images), or non-perceptual tasks, might better qualify as cognitive associations. (Parise, 2016, p. 23)

The above discussion is meant to illustrate that the two theory traditions used in this study are not at all incompatible. Furthermore, although my inclusive use of the term *cross-domain mapping* might come across as somewhat unconventional, it is neither unprecedented nor unjustified.

In conclusion, I will use the terms *metaphor* and *cross-domain mapping* as synonyms in this study. Both terms will refer to metaphorical expressions of musical harmony that utilize cross-domain mapping to conceptualize its perceptual qualities. In some cases, these mappings will be treated as indicative of an underlying conceptual metaphor, while in other cases they will be regarded as indicative of a cross-modal correspondence. There are two gains of this combinatory approach, in my view: Facilitating a broad investigation of cross-domain mappings in written text, unconstrained by theoretical dogma, while simultaneously allowing for a more nuanced discussion of the nature of individual mappings.

10 Method

10.1 Hypothesis

Because of the explorative nature of this study, I have made no predictions about the nature of cross-domain mapping of musical harmony in the corpus. However, the study was motivated by three hypotheses:

First, that literature on harmony *would* utilize such cross-domain mappings. This intuition was based on the long-term observation of such mappings in music (particularly jazz) vernacular and music-theoretical text, as well as on personal experience of metaphorical conceptualization of harmony. It has also been pointed out by Lawrence Zbikowski (2017b, p. 506), one of the leading theoreticians on music and metaphor, that almost all non-technical descriptions of music can be seen as having their basis in conceptual metaphors.

Second, that metaphorical mappings of harmony refer to a subjective perceptual reality. In other words, when authors describe specific harmonic phenomena as "bright," "thick," "sensuous," etc., these descriptions refer to a perceptual quality the author attributes to the harmonic stimulus.

Third, that metaphorical mappings arise because they reflect salient aspects of harmony perception. Therefore, even if certain metaphorical mappings may be completely personal and idiosyncratic, it is likely that intersubjective conceptual metaphors exist. These are potentially relevant to aural training insofar as they capture perceptual experiences of musical harmony that are shared by many people.

10.2 Qualitative document analysis

This study is a qualitative, explorative study of metaphorical expressions in a certain body of literature. As such, it falls into the category of *qualitative document analysis*. "Documents," in the present context, are texts that have been recorded without the researcher's intervention (Bowen, 2009), more specifically books about—or including discussions on—musical harmony.

Bowen (2009, p. 31) points out that compared to other qualitative methods, document analyses have both advantages and limitations. For the present study, relevant advantages include *availability* (all texts in the corpus are publicly available books), *lack of obtrusiveness* (the data is not influenced by being observed by the researcher), *stability* (documents are unchanging research objects), and *coverage* (the analysis can cover many different types of sources). The most relevant limitation is that of *insufficient detail*: Because the publications analyzed in this study were not created for the purpose of demonstrating cross-domain mapping of musical harmony, they may or may not contain examples of such mappings. Furthermore, since it is practically impossible to include *all* books about musical harmony in the corpus, the selection is susceptible to the researcher's biases. This threat was constrained by formulating clear sampling criteria. Finally, although Bowen mentions *efficiency* as an advantage of document studies, the opposite was in fact the case: Whereas Walther-Hansen's (2020) study was performed on a digital corpus of more than 50 million words, the present study relied on manual reading, making it a time-consuming method. This naturally placed a practical limit on the scope of the corpus, which was eventually set at 20 books.

The method used for analyzing metaphors will be described separately.

10.3 Sampling strategy

Given the vast number of available sources that could potentially be chosen for this study, some criteria of inclusion were formulated.

10.3.1 Published works

The internet has become an enormous source of music theory material, through blogs, podcasts, YouTube channels, Facebook groups, etc. While much of this content is amateur driven, there is also a significant amount of material created by professionals, for example musicians, composers, academics, or institutions. However, I chose to only include published books in this study. Partly because I wanted to focus exclusively on *written* language, and partly for reasons of verifiability.

10.3.2 Number of books

While a total of 20 books were included in the study, this number was not set in advance. Because the books varied enormously in length and use of metaphors, it was difficult to plan

exactly how many titles could be analyzed within the time frame set aside for this work. Though the number was arrived at by time restrictions rather than deliberate calculations, I will note that after approximately 10 titles I started to get a clear impression of the conceptual metaphors used across the literature. No additional categories were found in the latter half of the sources—only more examples of the already established ones. Therefore, it seems unlikely that a larger corpus would significantly alter the outcome of the study.

10.3.3 English language

While conceptual metaphors with similar structures may be found across different cultures, they are ultimately embedded in a specific language.

I realized at a very early stage that one of the most prevalent conceptual metaphors in the English sources, HARMONIC RICHNESS, does not appear naturally in the music vernacular of my native language, Norwegian. The metaphor would likely be *as understandable*, and with the same meaning as in English. But in my own experience (coming from a jazz background), a more idiomatic metaphor in Norwegian would be DISSONANCE IS FATNESS. A "rich" chord in English might rather be described as a "fat" chord (*en feit akkord*) in Norwegian (although possibly more often in jazz vernacular than in classical music).

While this may seem inconsequential, it became clear to me that comparing different languages could amount to a methodological challenge. For this reason, only books originally written in English were included in the corpus. This also excludes translations—an even more problematic category in this context since linguistic nuances may have been lost, distorted or "idiomized" in the translation.

However, this does not mean that the conceptual metaphors found in this study are only likely to be relevant to native English speakers. While a given conceptual metaphor may or may not occur naturally in a given language, there is nothing to suggest its structure cannot be *understood* across languages.

10.3.4 Variation and diversity

The most important consideration in the sampling strategy was to obtain maximum possible variation within a feasible number of sources. That is, the corpus was to include books that present the topic of harmony from a variety of positions, in a variety of contexts, and to a variety of readers. The corpus therefore includes different musical traditions (classical and contemporary classical music, jazz, pop, rock), different pedagogical purposes (aural training,

music theory, analysis, composition, songwriting, arranging, film scoring, music semiotics), and different model readers (readers with different levels of competence in music theory). The purpose of this variation was to ensure that the metaphors harvested were not all sourced from, for example, traditional studies on Eurocentric, eighteenth-century harmony, written in the 1950s. However, the corpus is still situated within a Western and educational context—as is this dissertation on the whole.

Furthermore, this study aims to examine a "slice of life" of how authors use metaphors to conceptualize harmony. Therefore, priority has been given to ensuring this particular slice has a little bit of everything in it, rather than choosing exclusively from the canon of treatises on harmony.[89] The purpose of the study is not to generate a predictive hypothesis about the metaphor usage of harmony treatises in general.[90] However, I believe diversity is the best way of ensuring the chosen works do not represent complete outliers in the general body of literature.

10.4 Analyzing the metaphors

The metaphors of interest in this study were those that in some way verbalize perceptual qualities of harmony by use of cross-domain mapping. This means that some metaphorical concepts fall outside the scope of this study, e.g., concepts like text painting or anaphones[91]. The same goes for metaphorical descriptions of a specific musical work such as an *agitated piece* or a *triumphant theme*. Finally, metaphors that are so fully integrated in music theory language that they have lost their *metaphoricity* (Gibbs, 2017) and become dead metaphors—e.g., *falling fifths*, *scale-step*, *parallel motion*—were not included in the analysis.

In the first part of the analysis, the books in the corpus were read manually, either on paper or on screen. Occurrences of harmonic metaphors were recorded and digitalized, along with the immediate context in which they appeared. In books covering a range of topics, sections less relevant to the subject of harmony were not analyzed (e.g., sections on music industry, studio production, film scripts, rhythm and meter, aleatoric music). To decide whether a word or phrase was metaphorical, the Pragglejaz Group's (2007) *metaphor identification procedure*

89 Besides, many of the authors in this canon—Schenker, Schoenberg, Hindemith, Rameau, C.P.E. Bach, or Zarlino, just to mention a few—wrote originally in other languages than English, and they thus fall outside the sample criteria set for this study.

90 Which is also the reason I will not make any systematic attempts at comparing the works based on genre, publication date, and so forth.

91 Anaphone: "using an existing model outside music to produce musical sounds resembling that model" (Tagg, 2004). See also *analogical reference* (Zbikowski, 2017b).

was used. This procedure contains a set of reliable criteria that are easy to use and produces results that are possibly falsifiable (Gibbs, 2017). The criteria are:

1. Read the entire text (i.e., written text or talk transcript) to establish a general understanding of the discourse.

2. Determine the lexical units in the text.

3. For each lexical unit in the text, check metaphorical use: Establish the meaning of the lexical unit in context (i.e., how it applies to an entity), and the relation in the situation evoked by the text (contextual meaning). You should take into account what words are before and after the lexical unit. Determine if the lexical unit has a more basic current/contemporary meaning in other contexts than the one in the given context. For our purposes, basic meanings tend to be: more concrete; what they evoke is easier to imagine, see, hear, feel, smell, and taste; related to bodily action; more precise (as opposed to vague); and historically older. Basic meanings are not necessarily the most frequent meaning of the lexical unit.

4. If the lexical unit has a more basic current/contemporary meaning in other contexts than the given context, decide whether the contextual meaning can be understood by comparison or contrast with the basic meaning. If yes, mark the lexical unit as metaphorical. Repeat steps 1–4 for each lexical unit. (Gibbs, 2017, pp. 69–70)

In the second part of the analysis, metaphors were assigned to different metaphor categories. These categories are not necessarily mutually exclusive, because words are often polysemous[92] and the use of a given metaphor is not always unambiguous. Take, for instance, the description "a somber chord." Using Merriam-Webster's definitions of "somber" (Merriam-Webster, n.d.-c), the expression may have several metaphorical meanings. It might indicate a conceptual domain of either SOUND-LIGHT ("somber: so shaded as to be dark and gloomy"), SOUND-COLOR ("somber: of a dull or heavy cast or shade; dark colored"), or SOUND-MOOD ("somber: of a dismal or depressing character; melancholy"). On the occasion that a passage did not offer enough context to distinguish between possible interpretations, more than one category coding was accepted. Therefore, a few words or phrases might appear several in the analysis, in different constellations.

92 Polysemy: the coexistence of several possible meanings for a word.

The digitalization, coding and analysis of the metaphors were done using NVivo (QSR International Pty Ltd., 2020).

10.5 Method of referencing

In the presentation of the results, I have made some choices regarding referencing for reasons of readability. In order to ensure textual flow, paraphrases are generally preferred over direct quotes. Numerous quotes consist of very short metaphors containing only a few words. Both paraphrases and short quotes are italicized, while quotation marks are reserved for longer direct quotes.

A single sentence in the presentation can frequently contain five or more quotations or paraphrases. Therefore, to avoid overloading the text, references are generally limited to specifying the source, without page number. Some quotes are very generic metaphorical descriptors that can be found in multiple sources (e.g., "colorful harmony"). When quoting such descriptions, I have chosen to give *no* reference. However, both source and page number for any quote/paraphrase in this presentation can be provided at request. The NVivo file—containing all citations, immediate context, page numbers, and coding—is also available at request by contacting me at villelangfeldt[at]gmail.com.

Lastly, in-text references do not contain the year of publication. No author is represented by more than one work in the corpus.[93] Therefore, it seems excessive to remind the reader that, for example, Persichetti's book was written in 1961, every time a reference is made to it. Rather, it will simply be referenced as (Persichetti).

10.6 Criticism of methodology

10.6.1 Limitations of document analyses

A qualitative document analysis is often—but not always—combined with other research methods in order to corroborate the findings, cf. triangulation (Bowen, 2009). However, the

93 Some authors are, however, represented by multiple works in the *dissertation*, and therefore have several listings in the final bibliography. In the context of the present analysis, references to such authors are always to the work that is included in the corpus (unless otherwise specified).

present study is an explorative one, whose main aim it is to *further the awareness and understanding of metaphors commonly applied to the description of harmony*. This does not amount to the confirmation or refutation of a hypothesis, or predictive claims about metaphors in books on musical harmony in general. Rather, the analysis is in itself the purpose: To provide empirical findings on how the perception of musical harmony may be conceptualized through language. Therefore, the research method is used here because of "its usefulness as a stand-alone method for specialised forms of qualitative research" (Bowen, 2009, p. 29).

10.6.2 Criticism of metaphors

There is substantial criticism of conceptual metaphor theory, both within and across academic disciplines, as thoroughly discussed in Raymond W. Gibbs' monograph *Metaphor wars: Conceptual metaphors in human life* (2017). Much of the "metaphor wars" are grounded in disagreement over what is fundamentally metaphorical, both in language and in human thought (Gibbs, 2017, p. 101). Another common criticism is that conceptual metaphors might not be more than post hoc artifacts *presented* as the fundaments of metaphorical thought (Gibbs, 2017, p. 12). Third, there is opposition to the entire idea that metaphors play any significant part in structuring human though, an opposition dubbed the *killjoy theory* by psychologist Steven Pinker (2007).

For the present study, these concerns are of little consequence. The claim made here is not that the perception of musical harmony is *based* on, or *contingent* on, metaphorical concepts. Instead, two weaker claims are made: First, I hypothesize that metaphorical expressions emerge *from* the attempt to verbalize auditory experience. In this regard, they are indeed "post hoc artifacts," but they are nonetheless relevant qua verbalizations of auditory experience. Second, I suggest that conceptual metaphors may still have the ability to *guide* auditory experience (cf. the "metaphorical" listening proposed in Part I of this dissertation). In other words, I propose that harmonic metaphors may simultaneously be a product of, and an influence on, auditory perception.

Another methodological consideration concerns the epistemological limitations of the metaphors themselves. Simply counting the instances of metaphorical expressions in a material is a rather coarse approach to capturing the nuances of metaphorical thought, holds Gibbs (2017, p. 94). Furthermore, even if metaphors succeed in accentuating particular perceptual aspects of musical harmony, these aspects encapsulate only a tiny part of musical experience. In his book on metaphor and musical thought, Michael Spitzer (2003) raises this concern:

> Metaphor theorists who assume that we conceptualize music in the same way we conceptualize language are naïve, or at least ill versed in aesthetics. The watchword of aesthetics is particularity, which means that art affords a richly grained mode of experience that is valuable precisely because it cannot be subsumed by concepts. (Spitzer, 2003, p. 77).

The present study does not claim that metaphors are able to capture all these dimensions of musical harmony. Rather, metaphors are regarded as conceptualizations which may highlight salient features of harmonic experience, thus potentially offering tools for holistic harmony listening in aural training education. Yet, the hypothesis—adopted from Walther-Hansen (2020)—that common metaphors in music arise because they resemble the auditory experience, may ultimately be misguided. However, this concern is best addressed through additional experimental studies. One such study will be presented in Part III of this dissertation.

11 Results

11.1 Short introduction to the thematic and metaphorical contents of each book

I begin by presenting the books that were chosen for the study, along with a short description of the contents of each of them. I also give a brief indication of how conceptual metaphors are used—or not used—in each work, so that it will become clear why some works are quoted heavily in this study, while others are never mentioned again. I have also made a note of the length of each book. The fact that the books range from 80 pages to 888 naturally has some impact on the number of metaphors sourced from each work.

The following books were selected for the study:

Jane Piper Clendinning and Elizabeth West Marvin (2016): ***The musician's guide to theory and analysis*****, 3rd ed.** (869 pages). A book written for a typical music theory course, with emphasis on the harmony of Western classical music, with elements of popular and contemporary music. Rich, varied, and relatively conscious use of metaphors.

Bill Dobbins (1986): ***Jazz arranging and composing. A linear approach*** (151 pages). A creative and detailed approach to arranging and composing for jazz ensembles. The *linear approach* to harmony is based on horizontal thinking rather than harmonic progressions and chord voicings. Metaphors are relatively few and conventional.

John M. Ferrara (2001): ***Jazz piano and harmony. An advanced guide*** (91 pages). Descriptive title. The book is short, and a large part of it consists of notated music. Very few metaphors were identified.

Gil Goldstein (1993): ***Jazz composer's companion*** (116 pages). A creative approach to jazz composition, based less on theory and more on concepts and the experience of other composers. Uses very few metaphors.

Ron Gorow (2006): ***Hearing and writing music: Professional training for today's musician*****, expanded 2nd ed.** (431 pages). A book that appears to be targeted towards studio musicians in the 1980s and 90s, covering everything from listening skills, harmony and arranging to publishing and music trade issues. Relatively few and conventional metaphors.

Dick Grove (1972): ***Arranging concepts complete: The ultimate arranging course for today's music*** (433 pages). Arranging techniques for big band or smaller ensembles in a jazz-oriented idiom. Grove uses two very specific and defined metaphors throughout the book: HARMONIC DENSITY (i.e., the number of different pitches being played simultaneously) and WEIGHT (i.e., the amount of doubling of pitches by the instruments involved). Very few other metaphors are used.

Mark Harrison (1995): ***Contemporary music theory (Level two). A complete harmony and theory method for the pop & jazz musician*** (380 pages). A theory book that seems targeted at a less academic audience. Harmony is presented in a simplified manner (compared to more conventional jazz theory), without necessarily resorting to a simpler harmony. Moderate use of metaphors.

Fred Karlin and Rayburn Wright (2004): ***On the track: A guide to contemporary film scoring.*** **2nd ed.** (532 pages). A comprehensive book covering most subjects relevant to film scoring. Harmony is only one of these but is treated both explicitly and indirectly through other subjects. The book contains interviews with film composers, so some of the metaphors may stem from interviewees rather than from the authors themselves. Moderate use of metaphors.

Gary S. Karpinski (2007): ***Manual for ear training and sight singing*** (398 pages). An aural training book from a central academic in the field. Very few metaphors identified, and mostly conventional ones.

Stefan Kostka, Dorothy Payne and Byron Almén (2018): ***Tonal harmony: With an introduction to post-tonal music*****, 8th ed.** (681 pages). Traditional approach to common-practice harmony and more contemporary classical music. Uses a broad variety of metaphors, although moderately, considering the book's length.

Steven G. Laitz (2008): ***The complete musician. An integrated approach to tonal theory, analysis, and listening*****, 2nd ed.** (888 pages). Thorough book on traditional Western harmony, with an emphasis on (Schenker-inspired) harmonic analysis. Moderate use of metaphors, considering it is the longest book in this study.

Mark Levine (1995): ***The jazz theory book*** (522 pages). An authoritative jazz theory book, written in a personal style and from an active musician's perspective. Relatively rich in metaphors.

Joe Mulholland and Tom Hojnacki (2013): ***The Berklee book of jazz harmony*** (264 pages). A jazz theory book covering both traditional and modern approaches to harmony. Rich and varied use of metaphors.

Ted Pease and Ken Pullig (2001): ***Modern jazz voicings: Arranging for small and medium ensembles*** (144 pages). Self-explanatory title. Emphasis on the vertical structure of harmony, seen through the lens of jazz theory. Relatively active use of metaphors, considering its modest length.

Jack Perricone (2018): ***Great songwriting techniques*** (400 pages). A book focusing on pop music, written for a less academic audience. Relatively rich and varied use of metaphors.

Vincent Persichetti (1961): ***Twentieth-century harmony: Creative aspects and practice*** (287 pages). The most metaphor-rich book in the study, in spite of its moderate length. One could open almost any random page and find a description of how a certain chord *sounds*. The variety is also impressive: Almost every metaphor category discovered in this study is represented in Persichetti's book.

Walter Piston (1987): ***Harmony*****, 5th ed.** (575 pages). A classic work on common-practice harmony. Written in a rather formal and dry style, but still with a certain quantity of metaphorical descriptions.

Paul Schmeling (2011): ***Berklee music theory book 2*** (80 pages). Jazz-oriented theory book. Only one metaphor was identified.

Philip Tagg (2013): ***Music's meanings: A modern musicology for non-musos*** (692 pages). A book covering an astonishingly broad spectrum of themes, from the perspective of music semiology. The parts dealing directly or indirectly with harmony have a moderate but interesting use of metaphors: Tagg explicitly discusses the concept of metaphorical perception of harmony, without necessarily using metaphors in doing so.[94]

Rupert Thackray (1993): ***Aural awakening: A course of aural training and general musicianship for students and teachers*** (277 pages). Rather old-school book on aural training. Not a single metaphor was identified.

94 Tagg's contributions to the subject of holistic harmony perception are discussed in Part I of the dissertation.

11.2 What is being mapped

Conceptual metaphors usually select certain aspects of the target domain. This means that other aspects may be obscured. For instance, TIME IS MONEY highlights that to each human being, time is a limited commodity which must be *spent* carefully. However, it does little to accentuate the vastness of time on a cosmic scale, or the relativity of subjectively experienced time. These aspects are ignored in TIME IS MONEY. What, then, is being mapped in cross-domain mapping of musical harmony—and what is obscured?

In the context of this study, the domain of MUSICAL HARMONY mainly contains *pitches*. Pitches are combined into CHORDS (pitch simultaneities) or CHORD PROGRESSIONS (successive pitch simultaneities), both of which form primary target entities for metaphorical mapping. In addition, the HARMONIC RELATIONSHIP between chords is also frequently targeted, especially in metaphors building in various ways on the conceptual domain of HARMONIC MOTION. Lastly, register (high/low) is occasionally part of mappings. It is seldom a target or source domain itself but may influence other conceptual domains indirectly.

Other features that could seem relevant to harmony—for example instrumentation, timbre, and to some degree even voicings—tend to be ignored in harmonic metaphors. This does not mean that these features are esthetically insignificant, but it may indicate that they are generally regarded as accidental rather than *invariant* properties of musical harmony. In Part I, using ecological theory, I made the claim that harmonic invariants can be traced to objective features of the auditory stimulus, namely *pitch structures* and their *relationship to the harmonic surroundings*. As shown here, the main contents of the MUSICAL HARMONY domain (i.e., CHORD/CHORD PROGRESSION and HARMONIC RELATIONSHIP) are therefore identical with the fundamentals of harmonic invariance. This suggests that harmonic metaphors, like harmonic invariants, may be rooted in objective auditory features—giving weight to the assumption that harmonic metaphors arise because they resemble the auditory experience.

11.3 Pitch-space metaphors

The idea of chords as *bodies of sound moving through space* is the basis for the conceptual metaphor PITCH-SPACE. Numerous conceptualizations of PITCH-SPACE are found in Western music theory, for example the circle of fifths, Schoenberg's (1954) charts of harmonic regions, models within a GTTM[95] framework (e.g., Lerdahl, 1988), and neo-Riemannian models of

95 *GTTM*: Generative Theory of Tonal Music (Lerdahl & Jackendoff, 1983).

Tonnetz, which draw on several of the aforementioned (Cohn, 2011). The metaphor is also a significant component of the music theory vernacular of both the English and other languages. Some of them are concepts without which it is hard to describe harmony at all. For example, consider the simple phrase "the dominant leads to the tonic." Here, two constellations of physical vibrations are first established as concrete entities (*the dominant* and *the tonic*). The first of them is then provided with agency and direction (*leads to*). When dissected like this, the metaphoricity of the phrase is obvious. Yet in everyday music terminology we do not really consider it particularly metaphoric that the dominant leads to the tonic. It is, arguably, a dead metaphor.

In this section I will analyze different manifestations of the PITCH-SPACE metaphor. While I have left out expressions I deemed dead metaphors, the mappings discussed here are still rather *conventional* metaphors. Some of them may seem obvious. Nonetheless, their importance and prevalence in the corpus justify their place in a study of this kind. Furthermore, Gibbs (2017) maintains that researchers "should not assume that conventional metaphors are never understood to express metaphorical meanings and do not count as instances of metaphorical language" (p. 67). Finally, explicating and naming these metaphors might inspire new pedagogical approaches in aural training, as I explore in the discussion. Precisely because of their commonness, they may be metaphors with a high degree of intersubjectivity and therefore easily "understood" by students.

11.3.1 Harmony has a physical body

The source domain BODY and the target domain HARMONY combine to form the conceptual domain of HARMONIC BODY. HARMONIC BODY can be said to be an ontological metaphor, one that establishes something abstract (in this case musical harmony) as an entity. Conceptualizing harmony as something physical facilitates speaking about what it *does* and how it is *constructed*.

The most central aspect of HARMONIC BODY is its conceptualization of harmony as something that takes up physical space. Persichetti in particular relies on this conceptualization: Chords are *bodies of sound* that are *surrounded* by other chords. They need *room for activity*, but risk *running into* each other or even *penetrating* one another. They may *expand* and *contract*, and can *increase* or *decrease* in size. Added tones are *attached* to the chord (all examples from Persichetti).

Both Grove and Harrison agree with Persichetti that adding tones to a chord makes it *larger*. However, the same idea is also conceptualized using WEIGHT and DENSITY respectively. Adding tones to a chord provides *weight* (Gorow, Persichetti) and can make harmonies *thick* (Laitz,

Persichetti). Grove introduces and explicitly defines the metaphor HARMONIC DENSITY as referring to the number of different pitches being played simultaneously: adding tones to a chord makes it *denser*. This definition is also used by Mulholland & Hojnacki, Harrison, and Persichetti respectively, although implicitly.

As with physical bodies, however, too much weight or density poses a risk:

> The increased harmonic weight … often presents the problem of immobility. (Persichetti, p. 81)

> [The chord] seems almost to sink under the weight of the dissonant altered tensions imposed upon it. (Mulholland & Hojnacki, p. 66)[96]

Too much density furthermore *reduces suppleness*, may *debilitate* the harmony, or even cause *harmonic suffocation* (all Persichetti). On the other hand, when chords or their texture are *lightened* or *thinned* by omitting tones, they become more *supple*, *agile*, or *flexible* (all Persichetti). If too many important chord tones are omitted, though, it can result in a *low-calorie* voicing (Mulholland & Hojnacki). Persichetti also describes the whole-tone scale as *meager* harmonic material (p. 55).

The WEIGHT metaphor is not used exclusively for denoting the number of pitches in a chord. Grove (and occasionally Persichetti) uses it to refer to the number of doubled notes in a voicing by the instruments involved. Thereby, doubling a triad with more instruments also adds WEIGHT to the harmony. Piston and Laitz use *weightiness* and *heavy* to describe the use of too many root-position harmonies, while Persichetti also associates it with register: "As a polychord is transposed upward it loses body" (p. 141). Finally, density and thickness may also result from a tighter chord spacing (Persichetti): "Voicings that contain all seconds create the maximum level of density" (Pease & Pullig, p. 93).

A handful of descriptors attribute harmony a PHYSICAL QUALITY, such as *sharp* (Persichetti), *soft* (Laitz; Piston; Persichetti), *soft-edged* (Piston), *hard* (Persichetti), *elastic* (Persichetti), *solid* (Mulholland & Hojnacki; Persichetti), *balanced/unbalanced* (Mulholland & Hojnacki), or *rough-hewn* (Persichetti).

In the preceding examples, HARMONY HAS A PHYSICAL BODY is used in a somewhat abstract sense, where the taking up of physical space is the central aspect. There are, however, two more elaborate versions of this conceptual metaphor, where the source domain is more concrete.

96 *Tensions*: Jazz terminology for *extensions*, i.e., pitches added to a triad or seventh chord.

The first builds on the *organic* body—or even the *human* body. It gives rise to metaphorical expressions that highlight a harmonic hierarchy: First, invertible counterpoint may become the *backbone* of the harmonic progression (Laitz). The most important harmonies then form the *skeleton* (Laitz, Persichetti), which may, finally, be *fleshed out* through harmonic elaboration (Laitz). The lack of such elaboration may expose a *bare-bones* sound (Perricone), whereas too much may cause *muscle-bound* chords (Persichetti). It should be noted that HARMONY HAS A HUMAN BODY is a clear influence in Schenkerian[97] analysis, both conceptually and linguistically. Therefore, it is not surprising that this conceptual metaphor is most visible in Laitz's book, as his is the most Schenker-inspired text in the corpus.

The second elaborated version of HARMONY HAS A PHYSICAL BODY specifies "body" as BUILDING. This metaphor connects to the conceptualization of musical composition—particularly of Western classical music in sonata form—as a type of ARCHITECTURE, where melody, motives, chords, and harmonic layout combine to form a well-designed construction. This can also be seen in traditional music theory's preoccupation with how chords are *built* or *constructed*. In the corpus, HARMONY IS A BUILDING is expressed in several ways: The bass line is the harmonic *foundation* (Karpinski). Lower pitches *support* higher pitches (Persichetti), or harmony *supports* the melody (Clendinning & Marvin; Pease & Pullig). Harmonic support consists of *harmonic pillars*, which in traditional harmony are identified by both Persichetti and Laitz as the tonic, dominant, and predominant[98] functions. Other functions *support* these pillars (Laitz). Clendinning and Marvin use HARMONIC PILLARS to conceptualize harmonic sequences and their performance:

> You could also think of sequences like suspension bridges with pillars at both ends holding up the bridge, but taut cables strung between them supporting the middle. In performance, focus on broad goals and motion between the pillars at the beginning and end. (Clendinning & Marvin, p. 407)

In more modern harmony, polyharmony can be created by *erecting* (Persichetti) *upper-structure* triads on top of *lower-structure* triads (Pease & Pullig; Mulholland & Hojnacki). Finally, chords and cadences can be *molded* into certain *shapes* (Persichetti). For example, "a pyramidal type of compound chord is composed of a series of intervals diminishing from the bottom upward" (Persichetti, p. 165).

97 Schenkerian analysis: A tradition that stems from the work of Austrian music theorist Heinrich Schenker (1868–1935). It involves exposing hierarchical structures of tonal harmony and has been highly influential on 20th century music theory, especially in the United States. See, e.g., Schenker (1969).

98 As Persichetti's book predates the emergence of the term "predominant" (i.e., a chord that normally resolves to the dominant), he specifies only the subdominant (Persichetti, 1961, p. 66).

11.3.2 Harmony is motion

The source domain PHYSICAL MOVEMENT and the target domain HARMONY combine to form the conceptual domain of HARMONIC MOTION. HARMONIC MOTION is a fundamental conceptual metaphor in Western classical music theory and is to a considerable extent incorporated in its very nomenclature: A succession of chords is a chord *progression*.

On the one hand, the conceptual domain of HARMONIC MOTION borrows content from tonal harmony, and influential theories like those of Rameau (1722/1971; see also Gur, 2008) and Riemann (e.g., 1886). These describe harmonic syntax in terms of motion, tension, and resolution: The dominant *leads* to the tonic, where the harmony *comes to a rest*, etc. On the other hand, it derives structure from the PITCH VERTICALITY metaphor, i.e., the concept of pitches ascending and descending through vertical space. The PITCH VERTICALITY metaphor gives rise to VOICE MOTION: a voice (or melody) goes up and down, rises and falls. It moves by *step*, *skip* or *leap*, or it can *travel*. Voices move in parallel, similar, contrary or oblique *motion*. The individual tones may be *leading* tones, *passing* tones, *escape* tones, etc. The metaphor underlying the movement of the single melodic line is expanded in the HARMONIC MOTION metaphor to describe the movement of entire chords or progressions, exemplified in this description by Goldstein (1993):

> Good voice leading involves the progression of one chord to another, smoothly, its notes moving up or down a half step, whole step, minor or major third, or remaining the same from one chord to the next. … Ideally, harmonic progressions should move like independent melodies. (p. 63)

HARMONY IS MOTION is so fundamental to the language of Western music theory that in its most basic form, as illustrated above, it falls outside the scope of this study. Yet, conventional music theory also incorporates other source domains to enrich the conceptual domain of HARMONIC MOTION. The most apparent one in the source material is that of TRAVEL. This mapping—creating the conceptual metaphor HARMONY IS A JOURNEY—allows for a whole structure of expressions describing different aspects of harmonic motion. Consider the following sentence:

> In this short excerpt [from the opening of Mahler's Symphony no. 9, fourth movement], what begins as a brief chromatic intensification of the home key (the modified sequence) turns into an excursion to remote keys, an excursion made smooth and orderly by a pair of chromatic mediant chains that lead eventually back to the home key. (Kostka et al., 2018, p. 449)

Here, Mahler's harmonic scheme is summed up lucidly in a single sentence, by help of the JOURNEY metaphor. We are given information about the direction (a "round trip" starting from and returning to the *home* key), the distance (*remote* keys), and the duration (*turns into*; *eventually*) of the tonal movement. The word "excursion" indicates that the aesthetic purpose of the journey is perhaps the shifting harmonic landscapes themselves, rather than an exact tonal destination. Furthermore, the "modes of transportation" give us insights into the harmonic techniques involved (chromatic intensification; chromatic mediant chains) and their esthesic effect (*smooth and orderly*).

Although prevalent in the study, metaphors built on A HARMONIC PROGRESSION IS A JOURNEY seem to be mostly reserved for tonal movements of a certain scope, i.e., progressions that incorporate actual modulation, or tonicizations at the very least. This is a perfectly logic expression of the JOURNEY domain: If you don't leave home (i.e., the home key), or you just visit *neighbor chords*, it hardly makes for an eventful odyssey. Thus, the sources offer many descriptions of what one could perhaps, tongue-in-cheek, call HOME SWEET HOME TONALITY. Here, we find chords that *have repose*, express *stability* or *immobility*, or *stop the harmonic flow* (all Persichetti). There is *static* harmony (Piston) with *a lack of direction* (Laitz), or music that *remains in the same harmonic sphere* (Persichetti). There are *paralyzed harmonic progressions* (Persichetti), *short modulations that move away and then quickly return to the original key* (Perricone), and chords that simply *expand the tonic area* (Clendinning & Marvin). Tagg's terms *chord shuttle* (oscillation between two chords) and *chord loop* (three or four chords repeated several times in succession) are also metaphors that describe a harmonic "commute" rather than a journey.

Rather, a HARMONIC JOURNEY is often initiated by a chord with a *restless* quality, or by harmonic *restlessness*. There must be *momentum*, *forward motion*, or *urgency*. Persichetti in particular repeatedly points to a kind of harmonic "urge": *a strong tendency to move on*; *clear directional powers*; *definite direction*; *strong directional pull*. When such elements are in place, a *departure from the tonic area* (Mulholland & Hojnacki) is possible, and the harmony may *move to a chord outside the established modal or key realm* (Persichetti).

In some cases, a HARMONIC JOURNEY makes use of a HARMONIC VESSEL—that is, a harmonic principle or technique by which a modulation (or series of modulations) is brought about. In the excerpt from Mahler's Ninth Symphony, mentioned earlier, this vessel was chromaticism and chromatic mediant chains. In the source material, chromaticism, secondary dominants, and leading-tone chords are generally deemed to be most effective in *intensifying the motion* toward new tonal areas and *creating forward momentum in a passage*. In other cases, harmony is not the vessel, but the very element through which one travels, like when *moving through*

a succession of chords (Persichetti). Other examples include *moving around the circle of fifths* (Mulholland & Hojnacki) or *traveling through a harmonic sequence* (Clendinning & Marvin). In quartal harmony of a certain complexity, *moving through inversions will produce real harmonic movement* (Persichetti).

Once a HARMONIC JOURNEY is under way, there is often an explorative element to it. Chords can be *adventurous* (Mulholland & Hojnacki), or the tonality may be *wandering* with *freedom to move from chord to chord* (Piston). Skillful composers are *able to guide their harmony in any direction* (Persichetti) and become "free to spread their wings and … explore new and often remote harmonic areas" (Laitz, pp. 757–758). However, explorative freedom need not be aimless, as careful harmonic planning may expedite either a distant or a sudden tonal relocation:

> Though surprising and quite **far afield** from C major, this D♭ **arrival** has been **prepared** by the ♭II (Neapolitan) in the recapitulation […]. (Clendinning & Marvin, p. 682, emphases added)

> In measure 2, V7/IV has ♭9 in the lead, **foreshadowing the upcoming** subdominant minor **incursion**. (Mulholland & Hojnacki, p. 121, emphases added)

In describing the structure of a fugue, Clendinning and Marvin depict the fugue subject as a sort of discoverer or settler:

> [An episode's] **harmonic goal** is typically the dominant (in major keys) or the relative major (minor keys). As the episode **prepares to establish the new key area**, several measures of free counterpoint may be added to lead to the cadence. **Once a new key has been reached, the subject appears** […]. (p. 500, emphases added)

Some HARMONIC JOURNEYS barely *touch on foreign harmonic territory* (Laitz), while others *travel through a variety of key areas* (Clendinning & Marvin). Extreme exploration might lead you to *the very frontier dividing tonal music from nontonal music* (Laitz).

A real odyssey must naturally contain obstacles along the way, and for a HARMONIC JOURNEY, *harmonic diversions* generally provide this drama. The harmonic progression might be *sidetracked* or *redirected* (Clendinning & Marvin). An expected resolution may be *evaded* (Clendinning & Marvin), or the *goal of a progression* ultimately *abandoned* (Persichetti). If one goes astray, one might "find refuge in a group of tones far removed from [the starting tonality]" (Persichetti, p. 250).

When a HARMONIC JOURNEY *veers into distant regions* (Laitz), one might stumble upon *foreign harmonic resources* (Persichetti) or *a dominant of the local key* (Clendinning & Marvin). Spending some time in a new key "allows the ear to get acclimated to the new tonal environment" (Laitz, p. 641).

In metaphorical expressions building on HARMONIC JOURNEY, what is primarily being mapped from the target domain of MUSICAL HARMONY is HARMONIC RELATIONSHIPS. HARMONIC JOURNEYS are strongly associated with harmony that is based on a hierarchy of "close" and "distant" chords or keys, and to a certain degree with the inference of a tonal center which constitutes the "home." Though this favors tonal music, many other types of harmony can also fit into the metaphor structure. However, it seems unlikely that HARMONIC JOURNEY would be an obvious conceptual mapping of for example dodecaphonic music.

11.4 Sensory metaphors

In this category, I present cross-domain mappings that combine the domain of HARMONY with the domains of COLOR or FLAVOR respectively. I have termed this category *sensory metaphors*. These are ways of conceptualizing the perception of harmony by linking it to other, more familiar sensory experiences: those of *seeing colors* and of *tasting*. Sensory metaphors, and in particular COLOR, are prevalent in the analyzed literature.

11.4.1 Color

This mapping combines the domains of HARMONY and COLOR to create the conceptual domain of HARMONIC COLOR. Versions of it are found in 14 of the 20 analyzed books, making it the one of the most common metaphor structures in the study. In some sources (e.g., Clendinning & Marvin), it is so prevalent that keeping count of its use is meaningless. Rather, I have focused on *how* it is used. Unlike most other conceptual domains in this study, HARMONIC COLOR is frequently specified in the literature—that is, authors explicitly use the term "harmonic color." Mulholland and Hojnacki also use a synonymous label at one point: *sonic color*.

Because of its ubiquity, HARMONIC COLOR may also be accused of being a dead metaphor. In its most basic form, like the actual term "harmonic color," this might indeed be argued. However, as I will show, the concept engenders a variety of metaphorical expressions, incorporating several different COLOR domains. Because of its productivity, I argue that the metaphoricity of HARMONIC COLOR is indeed intact. Furthermore, because the HARMONIC COLOR domain

is often used rather inadvertently, I believe it justifies a part in a study that aims to enhance awareness about harmonic metaphors.

While one might perhaps suspect HARMONIC COLOR to be a cross-modal correspondence—combining the modalities of hearing and vision—I argue that it is not. A cross-modal correspondence would suggest there are specific sets of correlation between the two modalities, for example the association of certain harmonic devices to certain colors. Leaving synesthesia aside, there is nothing to suggest such relationships between harmony and color perception on an intersubjective or universal level.[99] Neither do any of the sources in the corpus imply such relationships. Rather, HARMONIC COLOR appears to be a typical conceptual metaphor, structuring an abstract domain (harmonic experience) in terms of a more familiar domain (the experience of seeing colors).

However, some notable similarities between hearing harmony and seeing colors should be pointed out. First, musical harmony and color are both emergent features. While color is often thought of as an objective, physical property of the stimulus, Pomerantz and Cragin note that it is in fact a psychological one: "Wavelength is the corresponding physical feature, and color originates 'in the head,' from interactions of units that are sensitive to wavelength" (2015, p. 103). Second, there is an etymological connection: The Greek word for color, *chrôma*, is the root of the English "chromatic," which is used in connection with both musical harmony and colors. The meanings in the two domains are, furthermore, strikingly parallel, according to Merriam-Webster's (n.d.-a) definitions: First, *chromatic* relates to "the tones of the chromatic scale" and to "color phenomena or sensations" respectively. These primary, broad definitions are both contrasted with a more specific meaning: "characterized by frequent use of accidentals" and "highly colored" respectively. In other words, there is a linguistic parallel in the meanings of the word *chromatic* that is also reflected in how the conceptual metaphor HARMONIC COLOR is used in the corpus: Introducing *more* accidentals in the harmony leads to *more* color. I will return to this point later.

In the analyzed literature, the COLOR metaphor is used in two closely related yet distinct ways. In this section, I focus on the more basic and simpler version, namely HARMONY HAS COLOR. This version of the metaphor is rather neutral: It does not say anything about the *amount* or *strength* of color. It simply states that chords, scales, and so forth *have* particular "colors," which are different from the colors of other chords or scales:

99 Although in the writings of Messiaen, his highly personal color associations are oftentimes presented as if they were universal qualities embedded in the harmony itself: "In the gentle cascade of blue-orange chords … let us notice …" (Messiaen, 1990, p. 51).

The characteristic harmonic color of augmented sixth chords (Piston)

The color of the Neapolitan sixth (Piston)

Characteristic colors of the scale (Persichetti)

Colors of the primary chords (Persichetti)

The Phrygian color (Persichetti)

The function and color of the 13th chord (Gorow)

The tonal color (Perricone)

Dominant color (Mulholland & Hojnacki)

The basic color of a chord "is preserved [in inversions] but the overall sound of each inversion is unique" (Gorow, p. 239). Varying the harmony also changes its color, creating "a welcome relief from the prevailing harmonic color" (Laitz, p. 425) or a *change of chord color* (Clendinning & Marvin). A more discrete change can be achieved by adjusting one of the tones of a chord, creating a *color variant* (Persichetti).

I classify HARMONY HAS COLOR as the *qualitative* color metaphor. In this version of HARMONIC COLOR, "color" seems to refer to Gestalt qualities—distinct perceptual qualities that emerge from specific combinations of tones, without being present in any of the individual tones. A major chord, in this sense, will possess a "majorness color," while a minor subdominant has a "minor subdominant color," etc. The HARMONY HAS COLOR metaphor simply points out this perceptual quality.

Yet, the qualitative color metaphor is not as one-dimensional as one might think. In the corpus, it is significantly enriched by mapping expressions derived from the COLOR domain onto the harmony domain. A particularly common form of such mapping can be summed up as CREATING HARMONY IS LIKE PAINTING. It uses vocabulary associated with painting to describe various aspects of harmony: There is *basic color* (Gorow), *predominant*[100] *color* (Mulholland & Hojnacki), *pronounced color* (Piston), and *vibrant color* (Mulholland & Hojnacki). The aspiring composer must develop her harmonic *palette* (Clendinning & Marvin; Gorow; Kostka et al.; Persichetti) and practice at *creating a good harmonic blend* (Clendinning & Marvin). This

100 Predominant as in "most important," not the harmonic *predominant function*.

includes making *color modifications* (Persichetti), *accentuating* a color or increasing its *density* (both Persichetti), *introducing new color* (Piston), *vibrant color* (Mulholland & Hojnacki) or *fresh color* (Persichetti), *injecting a darker color* (Mulholland & Hojnacki) or *contrasting color* (Mulholland & Hojnacki), performing a *coloring of the harmonic surface* (Laitz), *coloring harmonies for contrast* (Laitz), or simply using a *variety of color* (Persichetti). This results in *nuance* (Laitz), *color contrast* (Laitz; Mulholland & Hojnacki), or *color gradations* (Persichetti). Persichetti also repeatedly distinguishes between *primary* and *secondary* chords. The primary chords have *primary color*. While he does not state it explicitly, he strongly implies that secondary chords, by the same token, have *secondary color*, e.g.: "The remaining chords are secondary and furnish variety once the primary color has been established" (Persichetti, p. 66). These metaphorical expressions seem to stem from the hierarchy of traditional color theory.

Failing to master the art of harmony blending, on the other hand, can lead to colors being *diffused* (Persichetti) or *muddy* (Schmeling; Persichetti), or the tonality may become *monochromatic* (Karlin & Wright). This, incidentally, highlights a weakness of sorts in the CREATING HARMONY IS LIKE PAINTING simile: In painting, a *tasteful use* of color is more important than *how many* colors are involved—Rembrandt and van Gogh are strikingly different in terms of how *much* color they use, but both are clearly masters of color. In the corpus, on the other hand, a *heavily colored* harmony (Clendinning & Marvin) or the *dramatic* use of color (Mulholland & Hojnacki; Laitz) is almost always viewed in a positive light. This tendency is even stronger in what I have analyzed as a second version of HARMONIC COLOR: the *quantitative* color metaphor.

The *quantitative* color metaphor, which I have termed HARMONIC COLORFULNESS, is more concerned with the "amount" of color provided by a harmonic device. Rather than using COLOR as a sensory metaphor to distinguish between different chords or tonalities (i.e., "this color is *different* from that color"), the HARMONIC COLORFULNESS metaphor suggests a hierarchy (i.e., "this chord is *more colorful* than that chord"). Harmony, according to this version of the metaphor, gets more colorful when more dissonance is introduced. In other words, color is something that is *added* to the harmony—through various techniques such as modal mixture and chromaticism.

The two versions are not independent from each other, and they clearly stem from the same conceptual domain of HARMONIC COLOR. Still, they are used in different contexts, and to describe different aspects of harmonic perception. In practice, HARMONIC COLORFULNESS is used synonymously with another large and important metaphor in this study: that of HARMONIC RICHNESS. For this reason, I will save the analysis of HARMONIC COLORFULNESS for the discussion on HARMONIC RICHNESS.

11.4.2 Flavor

This mapping combines the domains of HARMONY and TASTE to create the conceptual domain of HARMONIC FLAVOR. Its use is similar to that of HARMONY HAS COLOR, although it is less frequent. First, like the COLOR metaphor, HARMONIC FLAVOR is used to point out that different types of harmonic material *have* a certain "flavor," i.e., a distinct perceptual quality. Examples include, among others, *dominant flavor* (Laitz), *Mixolydian flavor* (Piston), *whole-tone flavor* (Pease & Pullig), *pentatonic flavor* (Persichetti), *polychordal flavor* (Persichetti), *quartal harmony flavor* (Karlin & Wright; Clendinning & Marvin), *blues* or *jazzy flavors* (Perricone), or *either flavor of a II chord* (Mulholland & Hojnacki).

Like with CREATING HARMONY IS LIKE PAINTING, the HARMONIC FLAVOR domain is further elaborated by incorporating expressions from a related activity: cooking. The metaphor CREATING HARMONY IS LIKE COOKING is reflected in expressions like *bring out the flavor* (Persichetti) or *making the flavor of each chord become more pronounced* (both Persichetti). *Mild* chords (Persichetti) and *bland* harmony (Ferrara, Persichetti) can be improved through *spicing* or *adding spice* to it in the form of dissonance or ornamentation (Persichetti). There is *fundamental flavor* (Persichetti), *nonessential flavor* (Kostka et al.), *piquant flavor* (Perricone), and *pungent flavor* (Piston; Perricone; Persichetti). A chord may become *acid* (Persichetti). Mulholland and Hojnacki note that hybrid chords[101] are a *low-calorie* voicing, missing that *tasty third* (p. 219).

11.5 Harmonic luminosity

11.5.1 Introduction

The concept of HARMONIC BRIGHTNESS/DARKNESS is not a novel one. As will be demonstrated, this metaphor is quite prevalent in the examined body of literature, although not distributed equally throughout the sources. It is a somewhat recognized concept in jazz circles and has been popularized in recent years by notable music theory YouTubers like Jacob Collier (e.g., Lee, 2017) and Adam Neely (e.g., 2016) in several videos. In contexts like these, the phenomenon is commonly explained with reference to the circle of fifths: "clockwise" harmonic movement creates a "brightening" effect, and "counterclockwise" harmonic movement creates a "darkening" effect (e.g., Lee, 2017). However, as the present study will demonstrate,

101 Hybrid chord: A triad played over a non-related root, e.g., a G major triad played over a bass C.

the concept of HARMONIC BRIGHTNESS/DARKNESS goes much further back and is decidedly present in Persichetti's 1961 book.

Rather than using the dual term HARMONIC BRIGHTNESS/DARKNESS (or just HARMONIC BRIGHTNESS, which it is sometimes called), I refer to brightening and darkening effects collectively as HARMONIC LUMINOSITY, following the example of Thoresen (2015). HARMONIC LUMINOSITY is one of many subtopics in *aural sonology*, a multi-method approach to analyzing music-as-heard developed at the Norwegian Academy of Music by Professor Emeritus Thoresen and colleagues.

Although the concept of HARMONIC LUMINOSITY is present in both popular culture, jazz vernacular, and musicology publications, no one—to the best of my knowledge—has ever conducted a comprehensive study on its history, its metaphorical structure, its phenomenological aspects, or—and above all—the specific musical mechanics behind its effects. This includes the aforementioned work by Thoresen, which is more listening-oriented than theoretically descriptive. The present study will offer a first step, by analyzing the occurrences of HARMONIC LUMINOSITY metaphors in the source material and explicating their specific harmonic context and content. This will lead to a rudimentary theory of the harmonic principles behind LUMINOSITY effects. However, this theory will be *entirely limited to descriptions given by the authors included in this study*. Sources on HARMONIC LUMINOSITY from outside the corpus will not be taken into account, nor will my personal opinions on the matter (except when explicitly flagged as such).

11.5.2 Results and discussion

There are different possibilities for classifying HARMONIC LUMINOSITY qua cross-domain mapping. I might have placed it together with the "sensory" metaphors of HARMONIC COLOR and HARMONIC FLAVOR since, like them, it appears to borrow perceptual experiences from a different sensory modality than hearing. This would indicate that experiences of brightening or darkening harmony are endo-musical qualities (i.e., existing completely "within" the music), but that we use a metaphorical language to verbalize them.

Without discarding this possibility, I propose that HARMONIC LUMINOSITY is indicative of a cross-modal correspondence. As such, it combines the modalities of *hearing* and *vision*. As discussed in the Theory chapter, I do not consider conceptual metaphors and cross-modal correspondences mutually exclusive, but rather as two different explanations of cross-domain phenomena, with different emphases and methodologies. Therefore, I analyze HARMONIC LUMINOSITY using a conceptual metaphor theory framework, but entertain the possibility that

it is based on integration of perceptual modalities. This is a somewhat stronger claim than defining it as a conceptual metaphor. The main reason for this choice is that unlike HARMONIC COLOR and HARMONIC FLAVOR, HARMONIC LUMINOSITY is used with surprising consistency. While there is no hierarchy of "chord flavors," or clear rules governing the amount of "color" in a chord, such structures are indeed implied in the metaphorical expressions stemming from HARMONIC LUMINOSITY. As part of the analysis of HARMONIC LUMINOSITY, I will therefore try to explicate these principles.

The question of whether HARMONIC LUMINOSITY is a conceptual metaphor or a cross-modal correspondence is only of theoretical significance in the present document analysis. It will nonetheless become relevant for the methodology of Part III of the dissertation, in which the perceptual salience of HARMONIC LUMINOSITY is tested empirically on 236 participants.

The analyzed literature is not always clear about exactly which source domain the "brightness/darkness" dichotomy stems from. Most often, it appears to refer to PHYSICAL LIGHT, giving rise to the conceptual domain of SOUND-LIGHT. This seems like the most plausible interpretation whenever the basic metaphors of *dark*/*darkening* and *bright*/*brightening* are used. Occasionally, SOUND-LIGHT is more clearly implied, although not named explicitly. For example, Mulholland and Hojnacki write that a V^{-7} (minor dominant seventh) chord is a darker replacement for a V^7 chord, and that using it in a tonic-dominant vamp will provide "an oscillation from dark to light" (p. 127). Here, the word *light* suggests a source domain of physical light. More often, however, the SOUND-LIGHT domain is implied through the use of expressions like *to foreshadow*/*foreshadowing*, *cast a shadow*, *carry a shading*, *somber*, *pale*, *brilliance*, *luster*, *shimmer*, and *sheen*.

In other cases, the terms *brightness* and *darkness* are explicitly linked to the source domain of COLOR, giving rise to the conceptual domain of SOUND-COLOR. This domain is expressed clearly by Mulholland and Hojnacki in variations over the metaphor *bright(er) color* (pp. 68, 71, 116), or *more* […] *colorful, whether brighter or darker* (p. 59), or to *inject a darker color* into a cadence (p. 122). Persichetti also expresses this when he refers to his dark-to-bright arrangement of the modes (see discussion on the "Dorian-centric" model later in this chapter) as a *color gradation* (p. 175).

Lastly, *bright* and *dark* are occasionally sourced from a MOOD or ATMOSPHERE domain, giving rise to the conceptual domain of SOUND-MOOD. SOUND-MOOD metaphors are primarily used to describe an esthesic effect, and typically link the words *bright* or *dark* to an affective descriptor, like *a brighter mood* (Mulholland & Hojnacki). *Bright* is coupled with descriptors like *positive* (Mulholland & Hojnacki), *open-ended* (Tagg), and *bright lift* (Mulholland & Hojnacki), while

dark is juxtaposed with *somber*, *sad*, *stark*, *brooding* (all Mulholland & Hojnacki), *ominous* (Tagg), *wistful* (Levine), and *heavy* (Persichetti). Mulholland and Hojnacki describe brightening modulations (i.e., clockwise in the circle of fifths) as *heightening the intensity*, whereas darkening modulations (i.e., counterclockwise in the circle of fifths) create a *relaxation of intensity* (pp. 166–168). The SOUND-MOOD interpretation has some support in empirical research: Eitan et al. (2012) suggest that cross-domain mappings may

> shed their sensory origins, as a domain originally serving as a target domain (emotion) becomes the new source domain. Thus, when sound is described as "dark," it may not be vision that directly serves as a source domain for a metaphorical description of sound, but "dark" emotion or mood. (p. 42)

The uncertainty associated with HARMONIC LUMINOSITY's specific source domain does not influence the metaphors' meanings. Lakoff and Johnson (1980, pp. 43–44) argue that related metaphors are more often *coherent* (i.e., they *fit* together) than they are *consistent* (i.e., forming a single image). In this sense, the two (or possibly three) versions of the LUMINOSITY metaphor are *coherent*, because the harmonic meanings they are ascribed are completely congruent: A given chord would not be labeled "bright" according to one version of the metaphor, and "dark" according to another.

The source material clearly treats HARMONIC LUMINOSITY as a continuum. It is not simply a brightening or darkening effect—the power of this effect can be assessed. Gradation descriptors found in the literature include *graduated*, *some*, *very*, *significant*, *distinctive*, *striking*, *dramatic*, *super-*, *extreme*, and *darkest possible*.

A total of 65 occurrences of LUMINOSITY metaphors have been identified and coded, across 9 of the 20 books in the corpus. However, they are far from evenly distributed across these 9 books:

Source:	**Occurrences of luminosity metaphors:**
Mulholland & Hojnacki	39
Levine	7
Clendinning & Marvin	7
Persichetti	6
Perricone	2
Ferrara	1
Karpinski	1
Kostka et al.	1
Tagg	1

Two or three of these occurrences are ambiguous; it is unclear if they actually refer to HARMONIC LUMINOSITY, or if descriptors like *bright*/*brightness* are mapped onto domains unrelated to harmony. The remainder of the occurrences clearly point to a perceptual quality of brightening or darkening *harmony*.

There seems to be a tendency to employ LUMINOSITY metaphors more often in literature dealing with harmonic principles that are freer than those typically associated with common-practice tonality. This tendency becomes even clearer when one looks behind the numbers of Clendinning and Marvin, the only common-practice oriented source with more than one occurrence: Six out of their seven LUMINOSITY metaphors applied are instances of the word *foreshadow* or *foreshadowing*. This "hallmark of Schubert's style" (p. 540) consists in "hinting" at a later modulation (typically from major to minor, or from major to a *darker* major key such as ♭VI) by inserting scale steps from the goal key into the tonic key (typically mixing scale steps like ♭$\hat{3}$ and ♭$\hat{6}$ into a major key). In other words, Clendinning and Marvin's use of LUMINOSITY is limited to a very specific metaphor, and a very specific harmonic context. Furthermore, the concept of *foreshadowing* is never explicitly connected to a concept of "darkening harmony," although this connection is systematically implied. Admittedly, it is contestable whether *foreshadowing* should be considered a LUMINOSITY metaphor, as its literal meaning is "to represent, indicate, or typify beforehand" (Merriam-Webster, n.d.-b) and has nothing to do with shadows. I have included it for three reasons. First, Clendinning and Marvin use it exclusively for depicting a *darkening* harmonic shift, thus implying a connection between harmonic effect and the word "shadowing." Second, the term is used by Mulholland and Hojnacki in the same way. Third, a have chosen to follow the strategy "When In Doubt, Leave It In," which recommends that it is better to mark a word as metaphorical whenever its metaphoricity is unclear (Gibbs, 2017, p. 72, attributed to Gerald Steen and colleagues).

A possible reason for LUMINOSITY metaphors appearing more often in books on jazz theory and twentieth century composition in this (albeit limited) study, is that these are musical styles with great harmonic freedom. In jazz, for instance, even the most basic II–V–I progression can be manipulated in numerous ways. Possibilities include dominantization, tritone substitution(s), chord extensions, non-diatonic added tones, or alterations (in addition to more conceptual approaches like slash chords and polychords). The musician is at liberty to use any of these techniques at will, and to give the harmony a personal touch. Adding a major ninth to a III^7 chord in a major key—for example, adding the tone F♯ to an E minor seventh chord in the key of C major (a brightening chord, according to Mulholland and Hojnacki)—is a perfectly acceptable artistic choice in most jazz contexts. However, the same solution would sound outlandish in the context of a Bach chorale, or a Haydn quartet. Although the harmonic styles of Bach and Haydn certainly include chromaticism, they adhere to other principles of

dissonance treatment, tonicization procedures, and rather strict notions of what is "inside" or "outside" the tonality at any given time. What passes as an inconspicuous "coloring" of a diatonic chord in a jazz tune, may not be a poïetic[102] possibility at all in common-practice harmony. In this freedom, one might speculate that jazz musicians are more inclined to letting their harmonic choices be guided by phenomenological awareness or esthesic effect, and, conversely, that they might be more attuned to perceiving and verbalizing such effects. In the same vein, one could argue that the greater harmonic freedom allowed by a musical style, the greater is the poïetic availability and relevance of LUMINOSITY effects. This is not the same as saying that the esthesic *effects* of LUMINOSITY are absent or less common in common-practice tonality—only that here, these effects are primarily governed by harmonic conventions, rather than by the personal taste of the composer or musician.

Levine's *The Jazz Theory Book* is a case in point. Levine uses both brightness and darkness metaphors, and quite explicitly points out their poïetic relevance: A Csus chord may be *darkened* by changing it to a $Csus^{(\flat 9)}$ (p. 322). Playing Dm^6 instead of Dm^7 as a minor tonic makes the chord sound *darker*[103]: "This doesn't mean you have to or even want to make this substitution. It just adds a different flavor" (p. 272). In an analysis of a chord progression by pianist Kenny Barron, Levine points out:

> Note the sudden darkening of the harmony when Kenny goes from the very bright A/G chord in bar 15 to $C-^{\Delta}/G$ in bar 16. Kenny uses the same effect on the final eight bars, alternating bright ($G^{\Delta 7}$) and dark ($C-^{\Delta}/G$) chords.[104] (p. 376)

Levine offers no explanations or musings about what, specifically, makes one chord brighter than another—he just asserts that it *is*.

Persichetti, too, highlights the poïetic possibilities of HARMONIC LUMINOSITY. Several places, he organizes harmonic material according to brightness and encourages the reader to use LUMINOSITY as a creative tool: "A workable order of the twelve ninth chords from darkest to brightest is material for a composer's craft" (p. 77).

In his chapter on scale materials, Persichetti also uses LUMINOSITY gradation as an ordering principle for the modes. A rather common pedagogical strategy is to present the modes by the order of their appearance when starting from Ionian and "moving upwards" (for example: C

102 *Poïetic*: Semiological term from Molino via Nattiez. Relating to the constructional aspects of music rather than its perceptional (cf. *esthesic*) aspects.

103 As a side note, Levine appears to be in disagreement with the other authors on this point. I will discuss this in more detail in Chapter 11.5.4.

104 The tonal context of the example is G major.

Ionian, D Dorian, E Phrygian, F Lydian, G Mixolydian, A Aeolian, B Locrian). The obvious drawback with this arrangement is that all modes now share the same accidentals and are prone to being perceived by the inexperienced learner as slightly different shadings of C major rather than as independent modes with distinct characters.

Persichetti instead presents the modes from a stationary key center, arranging them according to their brightness—defined by their gradually increasing number of sharps:

> The greatest number of flats that can be applied to a modal scale on a particular tone will produce the "darkest" mode, the locrian. Subtracting flats (and then adding sharps) in diatonic signature order will produce an arrangement of modes from "darkest" to "brightest." The dorian mode is the middle point and sets the norm. (p. 35)

He then goes on to demonstrate how this arrangement of the modes can be used creatively to create compositions that get increasingly brighter, by changing the *mode* rather than the key center.

This arrangement of the modes, which one could perhaps call *the Dorian-centric model*, can be found in a range of later publications (e.g., Brent, 2011; Mathieu, 1997; Sadai, 1980), including two publications in the corpus of this study (Kostka et al.; Mulholland & Hojnacki), and numerous online resources. To the best of my knowledge, however, Persichetti is the earliest source of this principle. In conclusion, although Persichetti does not use HARMONIC LUMINOSITY metaphors very *often* in his book, he uses them consciously and with specific intent.

As can be seen from the overview shown earlier, Mulholland and Hojnacki use the most HARMONIC LUMINOSITY metaphors in the corpus, with a total of 39 occurrences. This prevalence is not a coincidence: Like Persichetti, they use HARMONIC LUMINOSITY as a conscious tool. For instance, they adopt his Dorian-centric model of the modes, and they use LUMINOSITY to conceptualize the esthesic effects of modulation. And like Levine, they use LUMINOSITY metaphors actively as descriptors for the esthesic effect when judging between different harmonic choices. They regularly compare two possible chords for a harmonic situation on the basis of their LUMINOSITY:

> Compare the sound of figures 2.37 and 2.38: the B–7 chord in measure 2 of 2.38 is significantly brighter than the analogous [B–7$^{\flat 5}$] chord in 2.37. The F♯ creates a momentary distortion of the tonality of C major. But the strength of the root

> motion and essential voice leading still allow us to accept the chord as a brighter variation of VII–7♭5. (p. 53)

Like Clendinning and Marvin, Mulholland and Hojnacki use the FORESHADOWING metaphor. It is used with the same basic meaning: a diatonic chord is infused with non-diatonic material, to hint at a later resolution or modulation to a "darker" tonal area:

In measure 2, V7/IV has ♭9 in the lead, foreshadowing the upcoming subdominant minor incursion. (Mulholland & Hojnacki, p. 121)

> In this example, the related "II" chord in measure 2 has a different quality than we saw above: it is not –7, but –7♭5. This is a darker sound that effectively prepares the ear for the minor chord that is the ultimate target. (Mulholland & Hojnacki, p. 52)

Interestingly, they also describe an instance of *backwards* shadowing: "The final C– chord [in Tom Harrell's 'Sail Away'] casts a shadow over everything that has come before [which is all grounded in C major]" (p. 128). Furthermore, they describe *brightening* effects in the same vein as FORESHADOWING, although without using a specific counterpart term:

> As a creative choice, a minor 7 chord could be used in place of the [♯IV]–7♭5 in measure 6. The use of minor 7 as a related II is distinctive: It has a *brighter sound* because its 5 is now one half step higher than the tonic note. The increased chromaticism adds an *element of surprise and tonal ambiguity* to the progression. It *suggests resolution to a major chord*. (p. 55, emphases added)

As exemplified by this quote, Mulholland and Hojnacki often point to the specific harmonic details they believe are the cause the LUMINOSITY effect in question: In this case, the use of a perfect instead of diminished fifth in a ♯IV chord of a major key creates a raised scale degree Î, giving the chord a brighter sound. In addition, a ♯IV–7♭5 is the II in a II–V–I progression to the III– (minor) chord, whereas a ♯IV–7 is the II in a II–V–I to III (major).[105] Targeting a non-diatonic major III chord instead of a diatonic minor III– chord constitutes a brightening of the harmony.

In conclusion, Mulholland and Hojnacki use LUMINOSITY metaphors in a very conscious and systematic manner. They actively encourage the reader to think, listen, and harmonize

105 Or, in terms of classical music theory: The two versions of the ♯iv chord indicate tonicizations of different III chords: ♯iv7♭5 is a predominant targeting the diatonic iii (minor), while the ♯iv7 is a predominant to the chromatic mediant III (major).

in terms of brightness and darkness. They provide plenty of concrete examples of how to add LUMINOSITY to single chords, cadences, or progressions. Moreover, they differentiate and gradate the strength of LUMINOSITY effects in different situations. The intentionality of these metaphors can be seen in the nickname the authors give the ♭II$^{\Delta}$ chord[106], "the darkest of all the common modal interchange chords": *The Dark Lord* (p. 126). On using this chord as a replacement for a IV chord in a plagal cadence, they note: "Think of it as a *very* dark 'Amen cadence'" (p. 234).

To a certain degree, Mulholland and Hojnacki also discuss the underlying harmonic mechanics behind LUMINOSITY effects. However, these musings are generally limited to the specific LUMINOSITY instance in question, and rarely lead to formulations of a generic principle. Nonetheless, it is possible to induce some such principles from the rich examples offered by Mulholland and Hojnacki, along with the examples of other authors in the corpus. The last part of this section on HARMONIC LUMINOSITY is therefore an attempt at systematizing and generalizing the harmonic fundamentals that cause harmonic brightening or darkening, according to the source material in this study.

11.5.3 The harmonic fundamentals of luminosity

A detailed and thorough discussion of the specific harmonic mechanics that create LUMINOSITY effects are outside the scope of this study. Nonetheless, it is possible to present some rudimentary conclusions based on the use of LUMINOSITY metaphors found in the corpus. Since they each represent a concrete claim about a LUMINOSITY effect, it is possible to analyze the specifics of every individual harmonic situation, compare them to each other, and to extrapolate some general fundamental principles from them. These should be refined in future research, but they will provide sufficient basis for the quantitative study on metaphorical perception conducted in Part III of this dissertation. Furthermore, as will become clear, the different sources are relatively consistent and in agreement about what constitutes brightening and darkening harmonic effects. This indicates that a larger corpus might produce very similar results.[107]

LUMINOSITY is, above all, a relative phenomenon. The experience of *brightening* or *darkening* harmony is dependent on a harmonic context in which this change can take place and be perceived. A single chord is not "bright" or "dark," per se, it achieves this quality through juxtaposition or comparison with something else. The same goes for keys: Although listeners with absolute pitch might very well experience, say, E major as *inherently* brighter than A♭ major, this has nothing to do with HARMONIC LUMINOSITY as it is conceptualized in the

106 In the context of C major, this would be a D♭ major seventh chord.

107 Although in Part III I discuss an opposing theory of LUMINOSITY presented by Huron (2006).

corpus. It is the *relative* tonal movement or juxtaposition that creates brightening or darkening effects for instance in a modulation, no matter the precise tonal centers involved.

Nevertheless, I present the LUMINOSITY principles along two axes: First, I concentrate on a *contextual* dimension. This concerns luminosity differences between keys and modes and pertains to LUMINOSITY effects arising from shifts in tonality. Secondly, I look at a *single-chord* dimension, meaning the structure of—and LUMINOSITY differences between—single chords. I believe the contextual and single-chord dimensions are fundamentally not separable from each other, and they are expressions of the same basic LUMINOSITY principles. My reason for treating them separately in this presentation is clarity and the chance it provides to examine different *aspects* of LUMINOSITY.

As noted, a dissection of every instance of LUMINOSITY metaphors found in this study would warrant an article of its own. Therefore, I share the general reasoning used to arrive at the fundamentals of LUMINOSITY that I am about to present, rather than presenting every single quote that substantiates them.

11.5.4 Contextual luminosity

For the contextual dimension, the literature uses a clear principle for comparing the LUMINOSITY of one key or mode to another. Persichetti's Dorian-centric model of the modes is explicitly built on the premise that gradually *subtracting flats* and/or *adding sharps* makes the mode brighter—provided the root is kept stationary.[108] The relative LUMINOSITY of the different modes is supported by other authors in the corpus, either by reproduction of the entire Dorian-centric model, or by comparisons of two particular modes (Dorian is brighter than Aeolian; Aeolian is brighter than Phrygian; major key music is generally brighter than minor key music).

Extrapolating from this model, one arrives at a general hypothesis that exceeds the limitations of the seven common modes:

The more flats applied to (or sharps subtracted from) a scale, the darker. The more sharps added to (or flats subtracted from) a scale, the brighter.[109]

108 Exemplified from the root of D, this creates the following brightening order: D Locrian (three lowered intervals, or –3), D Phrygian (two lowered intervals, or –2), D Aeolian (one lowered interval, or –1), D Dorian (neutral, or 0), D Mixolydian (one raised interval, or +1), D Ionian (two raised intervals, or +2), D Lydian (three raised intervals, or +3).

109 Again, this is not to say that, e.g., D♭ major is inherently darker than C major because it has more flats. (Just like a "modulation" from D♭ to C♯, from five flats to seven sharps, would not constitute a massive brightening.) The relevant difference lies in the amount of raised or lowered intervals above the root, compared with another mode.

Although few sources explicitly address the LUMINOSITY of other modes than the seven common ones, there are a couple of examples in Mulholland and Hojnacki, who indicate that melodic minor[110] is brighter than Dorian (p. 94), and maintain that the altered dominant scale[111] "… is the darkest possible set of chromatic tensions that can be applied to the V7 chord; it represents a dominant chord with everything altered except for the essential chord tones $\hat{1}$, $\hat{3}$, and $\flat\hat{7}$" (p. 35). These examples support the hypothesis.

However, Levine offers a singular deviation from this principle. He maintains that "playing D–6 instead of D–7 [as a minor tonic] makes the chord sound darker" (p. 272). Let us analyze this claim. Playing a Dm^7 chord as a minor tonic would suggest a regular D minor key (i.e., D Aeolian, $\flat\hat{6}$ $\flat\hat{7}$), or possibly D Dorian ($\sharp\hat{6}$ $\flat\hat{7}$). Playing a Dm^6 would suggest D Dorian, or possibly D melodic minor ($\sharp\hat{6}$ $\sharp\hat{7}$). In neither of these cases would replacing a Dm^7 for a Dm^6 create a relative *darkening*, according to the principle hypothesized above. Of the three scales involved, Aeolian should be the darkest and melodic minor should be the brightest. I believe this anomaly stems from Levine having an impression of melodic minor harmony as being *particularly* dark—darker, even, than Dorian and Aeolian and all diatonic scales. He hints at this when he states that even though the only difference between a major scale and a melodic minor scale is the $\flat\hat{3}$, "melodic minor harmony *sounds* completely different—much darker and more exotic—than major scale harmony" (p. 57).[112] While it is not hard to agree that harmony built from melodic minor sounds strikingly different from harmony built from the major scale, it is an open question where melodic minor fits in the spectrum of HARMONIC LUMINOSITY. As was shown above, Mulholland and Hojnacki seem to disagree with Levine's sentiment, being of the opinion that melodic minor *is* brighter than Dorian. This is, in fact, the only place in the source material where two accounts of LUMINOSITY effects are incongruous. Based on the casual nature of Levine's use of LUMINOSITY metaphors—relative to the more systematic and coherent approach demonstrated by Mulholland and Hojnacki—I am inclined to regarding Levine as the outlier in this question.[113]

110 For readers not familiar with jazz theory, it should be noted that in jazz only the ascending form of the melodic minor scale is used, effectively making it a minor scale with $\sharp\hat{6}$ and $\sharp\hat{7}$ in both ascending and descending directions. It is used as a full mode rather than only in melodic contexts, with its own set of "melodic minor harmony." Thus, the only thing separating it from the Dorian mode is the $\sharp\hat{7}$, a relative brightness of $+\hat{1}$.

111 The altered dominant (or super-Locrian) scale is built from the seventh degree of the melodic minor scale.

112 It is clear—and also in accordance with the LUMINOSITY principles that are presented in this chapter—that *some* chords derived from the melodic minor scale are distinctly *dark*. Some examples include $sus^{(\flat 9)}$, which is darker than sus9 due to the flattened ninth, and V^7alt, which is darker than a regular V^7 chord (especially in a major key setting) in a range of possible voicings, due to it containing scale degrees like $\flat\hat{3}$, $\flat\hat{6}$ or $\flat\hat{7}$. One could speculate that the distinct darkness of individual melodic minor chords has led Levine to extrapolate this darkness to the whole scale, regardless of context.

113 A personal reflection: I absolutely agree with Levine that harmony derived from the melodic minor scale has a highly distinct *color*, that sets it aside from diatonic harmony. For me, melodic minor harmony has always evoked an association to a *cold-warm* quality—something that is cold and warm *at the same time*. This paradoxical quality can be illustrated with a situation well-known to many Norwegians: arriving at a freezing mountain cabin in winter,

The general flats/sharps principle we have formulated may also be conceptualized in another way: Clockwise movement in the circle of fifths creates harmonic brightness, while counterclockwise movement creates darkness. This conceptual model is in fact offered explicitly by Mulholland and Hojnacki (p. 167, see illustration), and is a common explanation for LUMINOSITY in less academic sources, for example, blogs and YouTube channels. Again, for clarity, the model is about LUMINOSITY effects arising from *relative movement*, not about the inherent LUMINOSITY of any specific key. C major is used as the center point in Mulholland and Hojnacki's model, but any key can act as the parent tonic.

Moving up by a fifth brings us to G major, which has one sharp more than C major. The next key, D major, has two sharps, A major has three, and so forth. Hence, a modulation from C major to G major creates a slight brightening of the harmony because one tone is raised (♯$\hat{4}$ of the parent tonic scale). A modulation from C major to A major, on the other hand, creates a more significant brightening, due to three tones being raised simultaneously (♯$\hat{1}$, ♯$\hat{4}$ and ♯$\hat{5}$ of the parent tonic scale). The opposite effect, a darkening of the harmony, is achieved by movement in the counterclockwise direction.

In effect, this spells out the same basic principle that we formulated above: *Adding sharps to (or subtracting flats from) a scale creates brightness*, and vice versa. The only difference is that this time we are not moving between different modes with a stationary root, but between identical modes with different roots.

As can be seen from Figure 10, Mulholland and Hojnacki argue that the LUMINOSITY effect is successively greater up to four steps along the circle of fifths. At the fifth step in the sharp direction, one arrives at the key a half step *below* the parent tonic. The modulation here will effectively create a darkening, due to every tone of the parent tonic scale being *flattened* by a half step. Going in the counterclockwise direction, the effect is mirrored: Every step is successively darker, until one reaches the fifth step, which constitutes a brightening modulation *up* a half step from the parent tonic.[114] At six steps along the circle of fifths one finds the tritone, which is neutral due to being both distant from the parent tonic and the meeting point of the brightening and darkening tendencies.

sitting by a crackling fireplace in full outdoor clothing, and waiting for the room to warm up. To me—for whatever reason—almost any chord of four or more tones from any degree of the melodic minor scale carries some of this quality. Perhaps this is reflective of the scale's luminous ambiguity: a relatively dark lower tetrachord ($\hat{1}$, $\hat{2}$, ♭$\hat{3}$, $\hat{4}$) fused with a relatively bright upper tetrachord ($\hat{5}$, ♮$\hat{6}$, ♮$\hat{7}$, $\hat{8}$).

114 Levine also describes a modulation up a half step as brightening (p. 396).

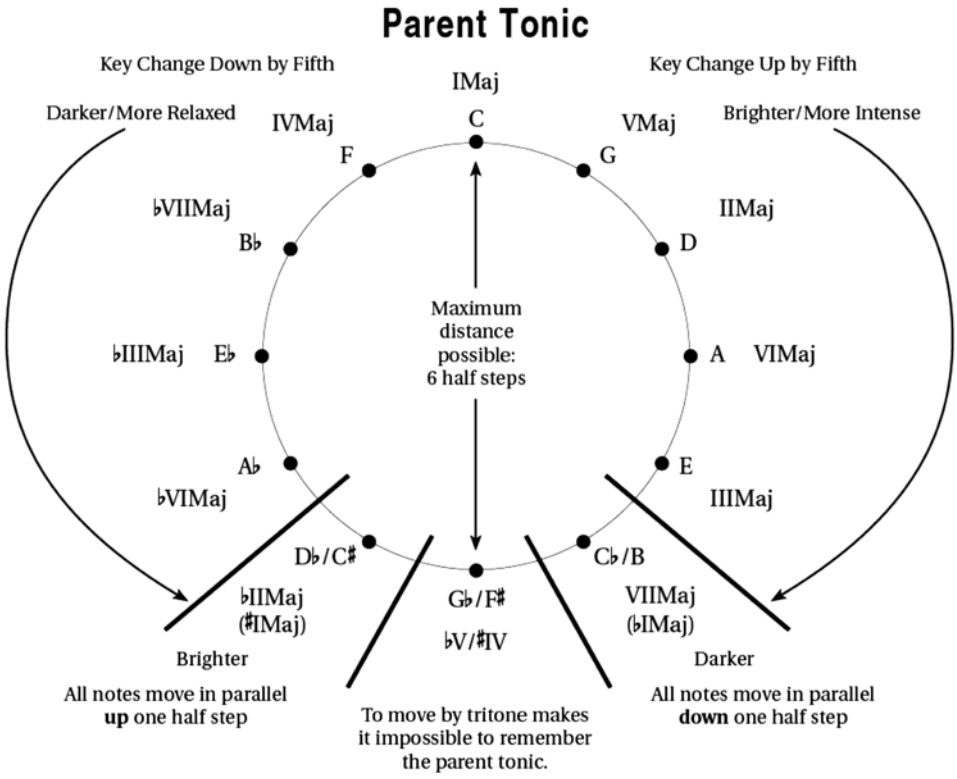


Figure 10: From Mulholland and Hojnacki (2013, p. 167): The esthetic effect of movement around the circle of fifths

The contextual dimension of HARMONIC LUMINOSITY can be summed up as adhering to the following core principle:

In general, major or augmented intervals above a mode's root create brightness, relative to minor or diminished intervals.

This principle can be conceptualized in two ways:

1. *Adding sharps (or subtracting flats) to a mode makes it brighter. Adding flats (or subtracting sharps) makes it darker.*

2. *Clockwise modulations along the circle of fifths generally create a brightening effect. Counterclockwise modulations along the circle of fifths generally create a darkening effect.*

11.5.5 Single-chord luminosity

The *single-chord dimension* of HARMONIC LUMINOSITY has to do with modifications of a single chord to create LUMINOSITY effects, that is, the principles of making a chord brighter or darker by changing its intervallic structure. As noted earlier, the contextual and single-chord dimensions are not two distinct phenomena, acting independently of each other. Rather, they appear to be two *aspects* of related harmonic mechanics—they interact and cannot, perhaps, always be distinguished from each other. One example of this is the single most common source of LUMINOSITY effects found in the analyzed literature: modal mixture (also known as modal interchange).

Modal mixture can be viewed as a principle by which two modes are "mixed"—hence the name. In practice, however, the effect is usually achieved by "borrowing" individual chords from another key. The various minor subdominant forms offer a particularly common source of mixture chords. Both Clendinning and Marvin and Mulholland and Hojnacki describe this effect—Clendinning and Marvin with their term *foreshadowing*; Mulholland and Hojnacki more specifically, by describing and comparing different subdominant mixture subdominants in terms of darkness: *darker*, *further darkening*, *very dark*, *darkest of all*. Mulholland and Hojnacki also offer several examples of *dark* or *foreshadowing* dominant chords—particularly in contexts of a major key, with altered dominant forms containing any of the scale degrees $\flat\hat{3}$, $\flat\hat{6}$, or $\flat\hat{7}$.

Brightening mixture is less common[115], but a few cases are mentioned in the corpus, although they are perhaps not examples of "real" modal mixture: Picardy thirds are brightening (Karpinski), II^7 can act as a brighter subdominant (Mulholland & Hojnacki), and the Dorian subdominant is a brighter alternative to the minor subdominant (Mulholland & Hojnacki).

As is apparent from these examples, LUMINOSITY effects of modal mixture adhere to the same core principle formulated under contextual LUMINOSITY: *In general, major or augmented intervals above a mode's root create brightness, relative to minor or diminished intervals.*

Other types of chromatic harmony mentioned also confirm the principle. Alteration of chord tones[116] generally adds LUMINOSITY to a chord, and the kind of LUMINOSITY gained is determined by the direction the chord tone is altered: Raised chord tones add brightness, flattened chord tones add darkness.

115 Modal mixture usually consists in importing elements of the minor mode to the major mode, not the other way around (Laitz, 2016, G-10).

116 Including chord extensions.

But a problem presents itself when a chord contains alterations of both kinds—both raised and flattened chord members. How does this influence the chord's LUMINOSITY? Do these tones cancel each other out, or are certain chord members more influential than others on the chord's overall brightness or darkness? While not addressing this issue explicitly, Persichetti (p. 77) hints at a sort of hierarchical order in his detailed presentation of ninth chords, arranged from darkest to brightest (Figure 11). In Persichetti's arrangement, all chords with minor ninths appear first, meaning they are the darkest. Then, all chords with major ninths, followed by chords with augmented ninths. Within these groups respectively, diminished sevenths come before minor sevenths, which again appear before major sevenths. After sevenths come fifths, still in ascending order, and finally thirds. Persichetti's rationale, whether thought out theoretically or discovered instinctively through listening, seems to be that the larger intervals override the smaller intervals: A ninth chord with a minor ninth is always darker than a ninth chord with a major ninth, no matter the composition of intervals below the ninth—and so on and so forth. It is naturally tempting to induce a general principle from this, claiming that the same hierarchy of intervals must apply to chord LUMINOSITY in general—e.g., seventh chords, or even thirteenth chords—not only ninth chords. This, however, would be pure speculation, as neither Persichetti nor any other author in the corpus present no such claim. Still, it might serve as a hypothesis for future research on LUMINOSITY.

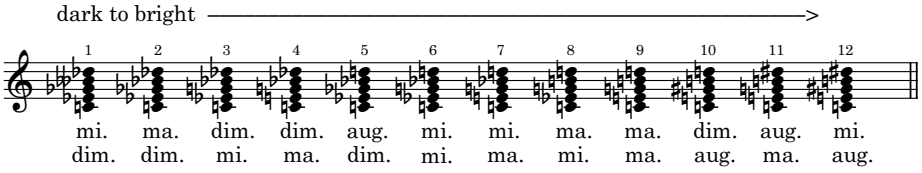


Figure 11: Persichetti's (1961) arrangement of ninth chords from darkest to brightest

In conclusion, the mechanics of single-chord LUMINOSITY—based on the analyzed material—can be summed up thus:

The chromatic sharpening of a diatonic chord tone or scale degree creates a brightening effect.

The chromatic flattening of a diatonic chord tone or scale degree creates a darkening effect.

If a chord contains both sharpened and flattened intervals, the overall luminosity of the chord is primarily determined by the highest interval, secondarily by the next-highest, etc.

As should be clear, this is in perfect accordance with the core principle of LUMINOSITY formulated earlier:

In general, major or augmented intervals above a mode's root create brightness, relative to minor or diminished intervals.

11.5.6 Concluding comments

The literature in this study is, for the large part, in agreement on the nature of HARMONIC LUMINOSITY. The authors who use LUMINOSITY metaphors use them in a consistent way, and generally agree on what sounds "dark" and what sounds "bright". The only instance of incongruity is on the status of melodic minor's luminous quality relative to the diatonic scales.

There is also one *apparent* incongruity that should perhaps be addressed briefly, to avoid any possible confusion: A modulation up a half step—e.g., from C major to D♭ major—constitutes a brightening, according to the principles I have formulated. Meanwhile, Mulholland and Hojnacki call the ♭II$^{\Delta}$ chord—e.g., D♭$^{\Delta}$ in C major—"The Dark Lord", and regard it the darkest of all common modal mixture chords. How can D♭ be both bright and dark, compared to C major? The answer is that it is not. In the first situation, D♭ is the target of a chromatically ascending modulation, in effect sharpening the whole key by a half step. In the other situation, D♭ is a replacement chord. The subdominant, F major, is replaced by F minor through modal mixture—a darker subdominant. The F minor is then, by the same token, replaced by its submediant D♭, a further darkening of the function. Thus, these two instances are not incongruous or in conflict. On the contrary, they exemplify the contextual and single-chord dimensions of LUMINOSITY respectively. However—to complicate matters even further—contextual and single-chord luminosity effects may occasionally be in conflict with each other. For example, a IIm$^{7(\flat5)}$ chord has a darker *structure* than "The Dark Lord," but the latter is darkest in a contextual sense. Which one of these two effects—single-chord or contextual LUMINOSITY—is the most salient? This might be dependent on the harmonic situation and the individual listener.

If one chooses to regard Levine's view on the melodic minor scale as an outlier, the analyzed literature systematically adheres to the core principle of intervals above the root. This principle constitutes a hypothesis that will be tested empirically in Part III of this dissertation. At the same time, it also represents a model for further theoretical studies. Despite the agreement on the fundamentals of HARMONIC LUMINOSITY found in this study, it was drawn from a limited corpus, and further research is necessary to get a better understanding of the phenomenon.

Lastly, I should point out a salient point concerning LUMINOSITY that has not been mentioned: Its relation to other aspects of acoustics. There are at least two aspects to be mentioned. One the one hand, Lasse Thoresen[117] and I have both made the personal observation that composers often intensify HARMONIC BRIGHTENING by combining harmony with other compositional techniques. Notable examples include the brightening of orchestral timbre, and the shifting of pitches towards a higher register.[118] Such effects are not discussed anywhere in the source literature, although Persichetti seems to hint at a natural connection between HARMONIC BRIGHTENING and accelerando, and between HARMONIC DARKENING and ritardando—at least, he proposes the *opposite* combinations as "provocative" creative possibilities (p. 277).

On the other hand—and this connects to first point—it seems obvious that a chord's timbral characteristics must also play a part in how its "brightness" is perceived. This should be problematized in future research on HARMONIC LUMINOSITY. The reason it is not discussed here is simply that none of the authors in the study make any comment on it. Also, in the greater context of aural training and holistic harmony perception, it seems perfectly defensible to claim that harmony has a *harmony-specific* luminosity component—e.g., that the major subdominant chord in the Dorian mode has a *brightness* that is unaffected by instrument, register, playing technique, etc.—and that this component can be examined and assessed regardless of timbral characteristics. And this is, in fact, the implied claim made in the literature in this study.

11.6 Harmonic richness

11.6.1 Introduction

The labelling of harmony or chords as "rich" is one of the most prevalent metaphorical descriptions found in this study. It combines the source domain of RICHNESS and the target domain of HARMONY to create the conceptual domain of HARMONIC RICHNESS. The general connotation of "richness" in this context is rather clear: The harmony or chord has "more" of something. However, what this "more" is, specifically, is seldomly explicated, making the conceptual domain of HARMONIC RICHNESS somewhat ambiguous. In the following presentation, I show how its source domain is frequently unclear, and how RICHNESS is occasionally replaced by other, synonymous descriptors. Apart from this conceptual complexity, HARMONIC RICHNESS is relatively straightforward. It prescribes relatively clear and gradated principles for creating

117 In private conversation.

118 Many lucid examples can be found in the orchestral music of Debussy and Ravel, for example in the second movement of Ravel's Piano Concerto in G major.

and comparing RICHNESS in different contexts. The last section of this presentation will try to explicate these principles.

11.6.2 Identification of source domains

Two basic categories of HARMONIC RICHNESS are expressed in the corpus. The first describes harmonic material (chord, progression) as *being* rich, while the second discusses how it can be *enriched*.

The first of these categories is rather one-dimensional. Similar to HARMONY HAS COLOR, it primarily points to a perceptual quality without specifying or quantifying it. Chords or harmonic materials are simply portrayed *as* "rich." There is, however, great variation in metaphorical descriptors. This reflects the multivalence of the word *rich*—it has several different, but related, meanings. Followingly, RICHNESS is expressed through a range of synonyms, implying different slightly different source domains.

Sometimes it is connected to a domain of LUXURIOUSNESS, shown in descriptors like *elaborated*, *ornamented*, *embellished*, *exotic*, or *lush* (multiple authors), *sumptuous* or *voluptuous* (Perricone), *sophisticated* (Harrison), *intense* (Karlin & Wright; Laitz), or *decorated* (Dobbins; Clendinning & Marvin). A common antonym is *simple* (multiple authors). Leaning on LUXURIOUSNESS, Grove notes that the richness of five-part voicings is *wasted* if the ear is not able to absorb it (p. 256). A few RICHNESS expressions possibly link to a TEXTURE domain, for example *thick*, *full* (multiple authors), and *saturated* (Harrison).

Most commonly, however, HARMONIC RICHNESS appears to be based on a somewhat surprising source domain: COLOR. This is most evident whenever the source is expressed directly, as in phrases like *add rich color* (Dobbins; Laitz; Perricone), e*nrich the harmonic color* (Laitz; Karlin & Wright; Persichetti), and *rich harmonic color* (Mulholland & Hojnacki; Piston; Perricone). Other places, the COLOR domain is expressed indirectly. This can be seen when expressions associated with RICHNESS/LUXURIOUSNESS and COLOR respectively are mapped onto HARMONY simultaneously: Persichetti speaks of *ornamental coloring* and *fringe color*. Clendinning and Marvin encourage *enriching* one's harmonic *palette*.

More importantly, the blending of the RICHNESS and COLOR domains in these examples is representative of a larger trend in the corpus: HARMONIC RICHNESS and HARMONIC COLORFULNESS are treated as synonyms. In the presentation of sensory metaphors, I only discussed HARMONIC COLOR used as a *qualitative* metaphor, i.e., used as a way of pointing out the

individual perceptual quality of this or that harmonic material. I will introduce the other version—HARMONIC COLOR used as a *quantitative* metaphor—at this point in the discussion.

11.6.3 Enrichment and colorfulness

The quantitative version of HARMONIC COLOR—HARMONIC COLORFULNESS—is less preoccupied with what the specific color *is* (e.g., "Lydian color" or "Neapolitan color"), than with the *amount* of color expressed in different harmonic materials. COLOR, in this version, is frequently complemented with a quantifier. In Mulholland and Hojnacki's book alone, the word *color* is coupled with the quantifiers *less*, *increased*, *more*, *a new level of*, *much*, and *tremendous amount*.

The reason I have introduced HARMONIC COLORFULNESS precisely at this stage is that it is used synonymously with, parallel to, and sometimes inseparably from, HARMONIC RICHNESS. Therefore, it makes sense to examine them in combination. The main purpose of these metaphors is not to describe something that *is*, but to convey *what can be done*. There is a prescriptive aspect to them, and they are most frequently used in combination with proposed manipulations of the harmony. Numerous metaphorical expressions describe how the harmony may be *enriched* or *added color to*, for example:

> [These examples] demonstrate how a standard basic progression in major may be chromatically **enriched** by including mixture chords. (Clendinning & Marvin, p. 535, emphasis added)

> The fourth is sometimes a useful note for minor chords, since it **adds harmonic color** without changing the basic minor sonority. (Dobbins, p. 13, emphasis added)

> Several of the extensions can be used together to **enrich** the basic chord structures. (Goldstein, p. 53, emphasis added)

> Any chord-tone may be altered by a half-tone to support a melodic tone or to **produce color** or tension. (Gorow, p. 260, emphasis added)

> This five-part form of the II chord is **a colorful option** in more advanced jazz styles, but would certainly not be used on a routine basis. (Harrison, p. 105, emphasis added)

> Polytonality can add intensity and luster by **enriching** the harmonic language and adding tension to a powerful statement. (Karlin & Wright, p. 233, emphasis added)

> These kinds of chromatic progressions—those that reinforce and **enrich** more basic diatonic progressions—are not the only ways that chromaticism can be employed. (Kostka et al., p. 441, emphasis added)

> We can **enrich the harmonic color** and smooth the linear motion of the sequence by using first-inversion chords. (Laitz, p. 548, emphasis added)

> To **add more color**, a tension can be stacked above a 3-note voicing. (Mulholland & Hojnacki, p. 212, emphasis added)

> We have already seen how a change of mode within the same key, from phrase to phrase, is a familiar **coloristic resource**. (Piston, p. 231, emphasis added)

HARMONIC RICHNESS and HARMONIC COLORFULNESS metaphors are found extensively throughout the corpus. Of the 20 books that were analyzed, only four did not contain any RICHNESS/COLORFULNESS examples: Those by Karpinski, Tagg, Schmeling, and Thackray. (However, across Schmeling and Thackray's books only one metaphor was recorded *in total*, so they are not only outliers in the RICHNESS/COLORFULNESS category but in the whole study.)

Their ubiquity makes HARMONIC RICHNESS and HARMONIC COLORFULNESS promising candidates for use in aural training. The potential is further strengthened by the fact that both metaphors prescribe clear principles for *how* to make harmony richer/more colorful. This is the subject of the next section.

11.6.4 Creating harmonic richness

What makes one chord sound richer than another chord, or one type of harmony richer than another type of harmony? In this section, I will try to formulate some generative principles of HARMONIC RICHNESS, based on the examples found in the corpus. In this discussion, I will rely on all synonyms for RICHNESS mentioned in the above presentation, including COLORFULNESS.

A large part of RICHNESS occurrences in the literature—unlike in the case of HARMONIC LUMINOSITY—are coupled with specific harmonic techniques that *cause* the richness. There are generally two aspects: The degree of RICHNESS in an individual chord or voicing (INHERENT RICHNESS), and the RICHNESS contained in a progression (CONTEXTUAL RICHNESS). These are not the same. The former strictly concerns a chord's structure, and thus determines its RICHNESS independently of the harmonic context in which it appears. The latter, CONTEXTUAL RICHNESS, is less one-dimensional, but frequently stems from the harmonic relationships

between chords. There is some overlap between INHERENT RICHNESS and CONTEXTUAL RICHNESS. For example, chromatically altering a chord may simultaneously enrich the chord and the progression it is part of. Yet the two aspects are based on partially independent principles. For example, a C major triad does not possess any significant INHERENT RICHNESS by any definition. Neither does it carry any CONTEXTUAL RICHNESS when it appears as the tonic chord of C major. However, if it appears as a secondary dominant in E♭ major—a chromatic chord—it brings CONTEXTUAL RICHNESS into the progression, while still having the same, negligible amount of INHERENT RICHNESS. This demonstrates that these two aspects of RICHNESS should be treated apart.

The inherent and contextual aspects can both be divided further into what I have termed the *quantity principle* and the *quality principle* respectively. Starting with INHERENT RICHNESS, the quantity principle dictates that an individual chord or voicing is enriched by adding tones to it. A seventh is the most elementary enrichment of a basic triad (Laitz; Persichetti; Perricone). Above the seventh, RICHNESS is further increased through ninths, elevenths and thirteenths (Clendinning & Marvin; Dobbins; Goldstein; Gorow; Harrison; Kostka et al.; Mulholland & Hojnacki; Pease & Pullig). Such tertian extensions, or *tensions* in the jazz vernacular, are generally added and combined more freely in jazz than in common practice harmony. As a result of this, some authors in the jazz tradition conceptualize chords simply by the number of their constituent tones (e.g., "five-part chords"), rather than by their specific tertian structure (e.g., "ninth chords"). Dobbins, Grove, Harrison, and Pease & Pullig all maintain that five-part chords are generally richer than four-part chords.

The additive methods mentioned thus far are, perhaps, rather obvious examples of HARMONIC ENRICHMENT—they all adhere to a "more-is-richer" principle. It should be noticed, however, that they can all be carried out using diatonic material. This is my reason for terming it the *quantity* principle: It demonstrates that HARMONIC RICHNESS can result from the quantity of tones alone, without employing chromaticism at all.

The quality principle, on the other hand, describes harmonic enrichment through *altering* chord tones. This principle is most active in CONTEXTUAL RICHNESS, through the use of chromaticism and modal mixture, but is also described as a RICHNESS factor in single chords. For example, Clendinning and Marvin (p. 601) and Pease and Pullig (p. 110) note that raising or lowering the fifth of a chord makes it richer, while Gorow (p. 260) maintains that this applies to *any* chord tone.

The quantity and quality principles can naturally be combined. For example, by altering a chord's extensions (♭9, ♯11, etc.) INHERENT RICHNESS ensues. Polychords[119] exemplify this:

> Polychord construction offers the conceptual advantage of combining two easily understood object to create something richer and more resonant. (Mulholland & Hojnacki, p. 215)

At the same time, polychords will enrich the harmonic context they are part of (Karlin & Wright), resulting also in CONTEXTUAL RICHNESS:

> Prolonged passages of pure polychords can result in tiring mass resonance; the richness and thickness stifle inner voices. (Persichetti, p. 147)

Progressions, too, can be enriched quantitatively, e.g., through introducing non-harmonic tones (Piston, 153). Generally, though, they are enriched *qualitatively*, through the introduction of chromaticism:

> Let's examine how traditional major/minor [tonal environments] **elaborated with chromaticism** create a **sumptuous** harmonic vocabulary. (Perricone, p. 112, emphases added)

> However, these kinds of **chromatic progressions**—those that reinforce and **enrich more basic diatonic progressions**—are not the only ways that chromaticism can be employed. (Kostka et al., p. 441, emphases added)

The chromaticism most commonly stems from altered chords (cf. INHERENT RICHNESS), applied chords (e.g., Laitz, p. 540), modal mixture (e.g., Clendinning & Marvin, p. 535), or bitonality/polytonality (e.g., Karlin & Wright, p. 267).

11.6.5 Does richness stem from dissonance?

I have shown that HARMONIC RICHNESS is primarily increased through the *addition* of tones (e.g., the quantity principle) or the *chromatic alteration* of tones (e.g., the quality principle). These principles beg the question of whether the fundamental and unifying principle governing HARMONIC RICHNESS is really *dissonance*. Is HARMONIC RICHNESS, then, in fact based

119 Polychord: A conceptualization of complex sonorities as a combination of two separate, less complex chords, e.g., "E♭m^7 *over* C^7."

on the conceptual metaphor DISSONANCE IS RICHNESS? In closing, I will briefly pursue this possibility. First, let us look at INHERENT RICHNESS.

The first indication that INHERENT RICHNESS is based on dissonance, is found in descriptions of intervals. Pease and Pullig note that the tritone is a *richer* interval than the fourth (p. 70), and also highlight the *richness* of the major seventh (p. 14). Though the DISSONANCE IS RICHNESS metaphor is generally only implied in the corpus, it is occasionally formulated more directly. In both of the following extracts, the dissonance produced by adding tones to a chord *causes* RICHNESS:

> Adding diatonic sevenths to triads does not change how these chords function, nor does it change root movements; it simply **adds dissonance**, **causing** additional momentum and **richness** to the progression. (Perricone, p. 112, emphases added)

> Tensions (9ths, 11ths, and 13ths) are the upper structure extensions of seventh chords. They are called tensions because they **create intervallic dissonance**. They **produce a richer**, more tense quality than that of a basic chord sound, which relies strictly on chord tones. (Pease & Pullig, p. 11, emphases added)

The same must by logical necessity be true for *all* instances where new tones are added to a chord: It will always make the chord more dissonant than it was in the first place. By the same token, suspensions must also increase the richness of a chord since they cause additional dissonance. This is supported by Clendinning and Marvin:

> In Example 16.25, from "Crazy for You" as sung by Madonna (1985), every chord is labeled "sus 2" or "sus," which gives the simple I-IV-V-I (E-A-B-E) chord progression a rich and lush sound. (p. 341)

It is also true that when chord tones are altered (cf. the quality principle), more dissonance is created. A chromatically altered chord is more dissonant than a diatonic chord containing the same chord tones (e.g., a $V^{7(\flat 9)}$ chord versus a V^{9} chord). We may thus conclude that the INHERENT RICHNESS of a chord is inseparable from its degree of dissonance. ENRICHING a chord necessarily makes it more dissonant. Therefore, INHERENT RICHNESS seems to be reducible to the metaphor DISSONANCE IS RICHNESS.[120]

120 On a side note, there must probably be some limitation as to how much dissonance would still be considered "rich," and at which point adding more dissonance would no longer make the chord "richer." This could, for instance, be the case with cluster chords. However, such limitations to HARMONIC RICHNESS are not problematized in any of the sources—the closest thing is a parenthesized remark by Mulholland and Hojnacki defining "color" as the

Moving on to CONTEXTUAL RICHNESS, the case is not as clear. The accumulation of dissonance is still a good explanation for many occurrences of CONTEXTUAL RICHNESS found in the corpus—the introduction of extended or altered chords in a progression, for example. However, many examples of CONTEXTUAL RICHNESS are achieved through harmonic relationships or juxtaposition, without the use of increased dissonance. The most notable example is modal mixture, which is frequently held up as an efficient way of creatin RICHNESS/COLORFULNESS:

> This passage is heavily colored by mixture chords, including the minor subdominant and tonic. (Clendinning & Marvin, p. 647)

Typical modal mixture chords highlighted for RICHNESS/COLORFULNESS effects are the minor subdominant in major (as in the above example) and the "Dorian" subdominant in minor (e.g., Mulholland & Hojnacki, p. 92). The Neapolitan is another chord that is often characterized as bringing RICHNESS to a progression:

> colorful Neapolitan harmonies (Clendinning & Marvin, p. 553)

> One of the more colorful chords that can be used to precede the dominant is the Neapolitan. (Kostka et al., p. 364)

Importantly, neither mixture subdominants nor the Neapolitan are by definition dissonant chords. They may all appear in the form of consonant triads—as may also applied dominants. This demonstrates that CONTEXTUAL RICHNESS, while often rooted in chromaticism, is not dependent on dissonance the same way as INHERENT RICHNESS. Rather, the RICHNESS may stem from the juxtaposition of diatonic and chromatic chords—or in other words, from harmonic relationships. Furthermore, some CONTEXTUAL RICHNESS descriptions simply rely on harmonic *variation*, without introducing dissonant elements:

> Although there are many examples of phrases built exclusively on tonics and dominants, most phrases incorporate the pre-dominant function to create a richer harmonic progression. (Laitz, p. 291)

> Notice the harmonic interest and richness in Example B compared with Example A, due in large part to the four different types of seventh chords used. (Laitz, p. 519)

"*stylistically acceptable* amount of dissonance" (p. ix, emphasis added), and the warning "use with caution!" about some particularly "colorful tension combinations" (p. 162).

In conclusion, DISSONANCE IS RICHNESS can be said to be the underlying metaphor for INHERENT RICHNESS (i.e., the RICHNESS of a single chord), but not for CONTEXTUAL RICHNESS. The latter may also stem from harmonic relationships.

11.7 Other metaphors

I have presented and discussed the most salient metaphors structures found in the descriptions of harmony in the corpus. Others were found but were so occasional that an individual presentation is not merited. Nevertheless, I conclude the analysis by outlining two categories that were still noticeable.

The first category is the use of personification. Chords or harmony are described as persons with human traits. The most obvious example is perhaps the already introduced ♭II$^{\Delta}$ chord in Mulholland and Hojnacki, *The Dark Lord*, which "appears in several guises" (p. 126). Another, no less chilling, character is the *Aeolian shuttle*, "a habitual harbinger of things dark and ominous" (Tagg, p. 341). More superficial descriptions include *wistful* (Levine), *gentle* (Harrison), *hopeful*, *yearning*, *brooding*, *bold*, and *assertive* (Mulholland & Hojnacki), and *insistent*, *stubborn*, and *overbearing* (Persichetti). One could perhaps suggest that these are expressions of the conceptual metaphor A CHORD IS A PERSON. On the other hand, they are rather contextual and do not seem to share any clear structures lent from the PERSON domain. Therefore, I am inclined to just calling them linguistic metaphors.

The second category is descriptions of mood or effect. Many descriptions fall into this category, like *humorous* (Dobbins), *mystical or dreamlike* (Karlin & Wright), *atmospheric* (Piston), *grandiose or heroic* (Karlin & Wright), *romance* (Karlin & Wright), *primitive mood* (Karpinski), *floating* (Levine; Tagg), *boring* (Levine), *happiness and calm* (Levine), *triumph* (Levine), *dark mood* or *darkness* (Mulholland & Hojnacki; Tagg; Levine; Perricone), *sad/sadness* (Mulholland & Hojnacki; Levine), *mystical* (Levine; Karlin & Wright), *ominous* (Tagg), *open-ended* (Tagg), *tense quality* (Pease & Pullig), *enchantment* (Levine), *urgency* (Mulholland & Hojnacki), *somber* effect (Mulholland & Hojnacki), *heightening/relaxation of intensity* (Mulholland & Hojnacki), and *feeling of closure* (Karpinski, p. 150). While descriptions like these may be just as relevant to aural training (cf. "metaphorical" listening) as conceptual metaphors, I have simply categorized them as *esthesic descriptors*, using Tagg's (2013) term.

12 Discussion

12.1 The role of metaphors in descriptions of harmony

My analysis of metaphorical language in the corpus has revealed that metaphors are used extensively to describe the perceptual effects of musical harmony. However, the distribution between sources is highly uneven. The outliers are the books of Persichetti (1961) and Mulholland and Hojnacki (2013), in which metaphorical expressions can be found almost on every page, on the one hand, and the books of Schmeling (2011) and Thackray (1993) which, combined, contain one single harmonic metaphor on the other. The other publications in the sample place themselves along a spectrum between these extremes.

One aim of the study was to search for common metaphor structures across different sources. Several such structures were identified. PITCH-SPACE metaphors were found in most works, variously depicting harmony as a physical entity moving through a virtual space. PITCH-SPACE is a significant factor in Western music theory, and a prerequisite in both traditional and contemporary conceptualizations of musical harmony (e.g., Rameau, Riemann, and neo-Riemannian theories). However, the analysis identified more elaborate versions of PITCH-SPACE model, such as HARMONY IS A JOURNEY. The HARMONY IS A JOURNEY metaphor maps vocabulary associated with the TRAVELLING, EXPLORATION, and TRANSPORTATION domains onto harmony, and is found across a substantial part of the works in the corpus.

A second common metaphor structure is that of HARMONIC COLOR. This is prevalent across the corpus and is used in two ways. It may describe the particular perceptual effect of a particular chord or harmonic material (e.g., *Lydian color*). More generally, it is used synonymously with HARMONIC RICHNESS: to conceptualize how harmony may be elaborated (i.e., *enriched* or *colored*) through various means.

A third metaphor type is HARMONIC LUMINOSITY, which describes harmonic effects in terms of *brightening* or *darkening*. LUMINOSITY metaphors are used extensively by a few authors—in particular Mulholland and Hojnacki, who even present a theory about modulatory processes and their LUMINOSITY effects. These metaphors are not prevalent across all sources in the corpus, however. I have presented the possibility that HARMONIC LUMINOSITY should not simply be regarded a conceptual metaphor but may in fact be indicative of a cross-modal correspondence between the auditory and visual domains. While this claim cannot be assessed

within the methodological framework of a document analysis, it may be addressed by future research. One attempt is found in Part III of this thesis.

What types of harmonic material or situations are most prone to being described metaphorically? Based on the findings in this corpus, a general hypothesis may be formulated: *The more harmony deviates from strict diatonicism, the more likely it is described metaphorically*. This means that metaphorical expressions are most often (but not exclusively) used in describing the aural effects of modal mixture, chromaticism, tonicizations, and modulations.

Lakoff and Johnson (1980) define *coherent* metaphors as *fitting together*, and they argue that metaphors are more often coherent than they are *consistent*, that is, forming a single image (pp. 43–44). This is also the clear tendency found in the present study: Metaphors found across the sources were relatively coherent. While the conceptual domains of for example HARMONIC LUMINOSITY and HARMONIC RICHNESS are ambiguous and thus engender a variation of metaphorical expressions, the expressions are relatively coherent with each other. The metaphors of HARMONIC LUMINOSITY and HARMONIC RICHNESS were analyzed in detail to elucidate the harmonic principles governing them. These analyses supported the impression of coherence, although HARMONIC RICHNESS was found to be slightly less coherent than HARMONIC LUMINOSITY.

12.2 Relevance for aural training

What does it tell us that certain authors of books on harmony use a lot of metaphors, while others do not? A central hypothesis in this dissertation is that metaphorical listening is a sort of Gestalt listening. A description of a chord's *color* or its degree of *richness*; a notion of the gradual *brightening* of a progression—these are all descriptive of an "overall feel," or, rather, of a Gestalt quality. As such, metaphors may play an active role in the development of Gestalt listening skills, by helping students notice and verbalize what they hear. Harmonic metaphors are particularly promising in this respect if they are shown to be perceptually salient for large groups of people. That is, if there is empirical and significant agreement on issues like when harmony sounds "brightening" and when it sounds "darkening."

Naturally, some metaphors are more relevant to a pedagogical context than others. But any good metaphor can have a place in an aural skills teacher's repertoire. The basic sensory metaphor HARMONY HAS FLAVOR has a rather one-dimensional structure. It differentiates between "flavors," but has no comprehensive ordering principles for *comparing* these flavors to

one another, other than applying adjectives from the TASTE domain (e.g., *sour*, *spicy*). Yet, the FLAVOR metaphor can be put to good use: Encouraging students to notice a chord's particular *flavor* is a quick and likely much more intuitive approach than asking them to focus on its *Gestalt quality*. This offers further possibilities, as a way of helping students become aware of their *own* associations to the chord: *Listen to this V7/V chord. Notice its characteristic flavor. How would* you *describe this?*

This brings us to what I believe is the most important point in this discussion: The difference between an *inadvertent* and an *explicit* use of metaphors. Although it merely speculation, I will venture to say that the majority of metaphors found in the analyzed literature are inadvertent metaphors. At the very least, it is not often clear whether the author uses them actively to draw attention to a particular aspect of harmonic experience, or just as a figure of speech. From the viewpoint of conceptual metaphor theory, this is not important: Most of the conceptual metaphors we rely on in everyday life are used more or less without awareness of their metaphoricity (Lakoff & Johnson, 1980). In a pedagogical context, however, I believe this makes all the difference. Consider the difference between these two (hypothetical) approaches to describing the use and effect of modal mixture:

1. "Modal mixture can be used to color diatonicism."

2. "Imagine harmony has *color*. Which of these two cadences would you say is the most *colorful*? Which specific chord do you think adds the most color? How would you describe this chord theoretically, and what is the reason it brings a new color into the progression?"

The first approach uses HARMONY HAS COLOR, but does nothing to explicate either the metaphor itself or its harmonic meaning. Here, the metaphor embedded in the sentence may just as likely pass unnoticed by the student. The other approach highlights the metaphor, and lets the student discover its implication herself, through perceptual experience.

There are, naturally, elegant cases of explicitly used metaphors to be found in the corpus. In this example, Clendinning and Marvin builds on HARMONY IS A JOURNEY to explain harmonic sequences:

> Sequences are like subways, buses, or trains: you get on at one location, ride for a while, and get off when you reach your destination. … Some sequences "travel" within one key area, while others move from one key to another. (2016, p. 407)

This metaphor communicates two important facts about harmonic sequences: First, they are flexible and do not necessarily have fixed starting points and harmonic goals. Second, sequences can be either modulatory or non-modulatory. This may naturally be conveyed through less metaphorical descriptions, as by Kostka et al.: "A **tonal sequence** will keep the pattern in a single key, which means that modifiers of the intervals (major, minor, and so on) will probably change … A **real sequence** … transposes the pattern to a new key" (p. 96, emphases in original). These extracts are admittedly taken out of their context. Yet, they illustrate saliently how meaning and intention can be adjusted through an explicitly used metaphor. Clendinning and Marvin's description points to perceptual *effects* of harmonic sequences (i.e., a feeling of harmonic movement). The description of Kostka et al., on the other hand, has a more constructional emphasis (i.e., whether intervals are modified or not). While it does utilize some dead metaphors (*keep in* a key; *transpose to* a new key), these do nothing to verbalize esthesic aspects of harmonic sequences.

12.3 Summary and criticism

Through the "metaphorical" listening approach introduced in Part I, the listener focuses attention on emergent features of musical harmony. Hearing a chord or progression as "being" spy-like, bright, warm, acid, longing, rich, etc., is suggested as a way of noticing its holistic quality. While some associations may be restricted to an individual listener, I have proposed that more intersubjective metaphors likely exist in the vernacular of Western music theory. The present study supports this assumption. Though based on a limited corpus of 20 written works, the document analysis shows that the majority of metaphorical expressions adhere to common conceptual metaphors. Moreover, the metaphors are to a substantial degree used coherently in unrelated works. This gives weight to the claim of intersubjectivity and may furthermore be taken as support for the assumption that harmonic metaphors arise because they resemble the auditory experience (cf. Walther-Hansen, 2020). Nonetheless, the possibility should not be overlooked that typical metaphor structures are spread from author to author through exposure to an established music-theory vernacular. Potentially intersubjective metaphors are particularly interesting in light of aural training, because they may show to be perceptually "understandable" to a large number of listeners. The implication is that such metaphors can be utilized in aural training to verbalize perceptual qualities of harmony, thereby preparing the ground for holistic harmony identification skills.

It may be objected that psychoacoustic effects of timbre, instrumentation, texture, etc. are ignored in the discussion of metaphors. However, the analysis uncovered that only the

(so-proposed) harmonic invariants[121] of *pitch structures* and *harmonic relationships* tend to be mapped in harmonic metaphors. Other constructional aspects of harmony were largely ignored in cross-domain mapping in the analyzed literature.

A second possible objection is that the role of individual differences, musical training, social context, and other themes are not considered in this study. These are relevant effects on the perception of musical harmony that certainly warrant closer inspection. However, they fall outside the aim, scope, and methodological framework of a qualitative document analysis. They are better assessed through other methodological approaches, something which should be done in future research. For example: Even assuming consensus on the tonal principles behind HARMONIC LUMINOSITY, are listeners generally able to *perceive* LUMINOSITY effects, and do they perceive it in a homogenous way, in accordance with the principles? If they do, this opens up interesting new questions that are especially pertinent to aural training purposes: Is the pick-up of LUMINOSITY effects a universal skill, or characteristic of specific populations? Is it acquired through general enculturation, or through the acquisition of musical ability? Are there differences in perceptual skills between musicians from different genres, such as classical music and jazz, or between musicians with different instruments, such as melodic instruments and chord instruments?

Because of the possible untapped potential for holistic listening approaches represented by HARMONIC LUMINOSITY, I have chosen it as the subject for the quantitative study that forms Part III of the dissertation.

121 See Part I for a discussion of perceptual invariance.

PART III

Brightening chords and darkening progressions.
A quantitative study on the perceptual salience of harmonic luminosity in musician and non-musician populations

13 Introduction

13.1 Background

Earlier in this dissertation, a holistic listening approach to harmony was defined as directing one's attention towards a chord's or chord progression's overall quality—its emergent features—rather than towards any of its constituent parts. The emergent features of harmony generally appear to possess metaphorical qualities: Chords or chord progressions are experienced as "being" or "being like" something. Such metaphorical qualities may fall into a range of non-mutually exclusive categories, for example pertaining to harmonic structure or psychoacoustic effects (e.g., *mellow*, *energetic*, *dense*, *hollow*, *clustered*, or *expanding*); stylistic or cultural-semiotic connotations (e.g., *blue*, *jazzy*, *Messiaenesque*, or *spy chord*); cross-modal descriptors (e.g., *sharp*, *warm*, *cool*, *bright*, or *dark*); affective descriptors (e.g., *happy*, *sad*, *bittersweet*, or *longing*); or functional descriptions (e.g., *cadential*, *centric*, *chromatic*, *doo-wop progression*, or *four-chord song*).

The ability to identify harmonic material as Gestalts is considered an expert skill, but aural training teaching methods have traditionally lacked strategies to develop this skill deliberately (Karpinski, 2000). In the present dissertation, practicing recognizing and labelling metaphorical, emergent qualities in chords and chord progressions is proposed as one concrete strategy. However, some such qualities may be strictly personal associations (e.g., that a certain chord sounds "Christmassy"), while others might conceivably be cultural or even cross-cultural. Whereas both personal and cultural metaphors may play an important role in an individual's development of Gestalt listening skills, the latter make for more hands-on methodological possibilities in aural training class settings.

The experiment presented here examined the perceptual salience of a specific category of cross-modal correspondence: *harmonic luminosity*.[122] This category was chosen on the basis of a qualitative document analysis examining the use of conceptual metaphors and cross-modal correspondences in a selection of literature on harmony (see Part II of this dissertation). Harmonic luminosity is a purportedly systematic association between the perception of musical harmony and the idea of visual brightness.[123] The label *harmonic luminosity* was

122 In Part II of this thesis, "harmonic luminosity" was consistently written in small caps (HARMONIC LUMINOSITY), according to custom in conceptual metaphor theory. In the present chapter, harmonic luminosity will be treated as a cross-modal correspondence and will be written using a normal font.

123 Note that this is not a form of synesthesia. In synesthesia, perceptual input in one modality triggers sensations in another, unrelated modality. Harmonic luminosity, rather, is portrayed as a phenomenon where certain types

coined by Thoresen (2015), and is not, therefore, a universally established term. The concept it represents, however, was found relatively frequently in harmony textbooks back to 1961 (Persichetti, 1961) in the document analysis in Part II. In a few cases the concept was explicitly explained, in other cases implied through metaphorical language. While the present dissertation does not include a review of harmonic luminosity metaphors in music theory literature prior to 1961, the concept can be found at least as early as Riemann. In his *Musik-Lexikon*, Riemann writes the following under the headword "Variante"[124]:

> **Variante**, ein vom Verfasser d. L. erst in den letzten Jahren in seinen Schriften eingeführter Terminus, der die durch Veränderung der Terz (groß statt klein, klein statt groß) substituierte Durform der Tonika der Molltonart oder die Mollform derjenigen der Durtonart kennzeichnet. Der Ausdruck ist deshalb gewählt, weil bei solcher Substitution gewöhnlich keine eigentliche Modulation stattfindet und mehr nur ein plötzliches Heller- bezw. Dunkler-werden der bleibenden Tonart vorliegt.[125] (Riemann, 1900)

Harmonic luminosity can also be found in academic discourse such as Thoresen's monograph (2015), as well as his unpublished work on the subject (including analyses of luminosity effects in classical pieces for educational use). Finally, the concept has been popularized in recent years in online music theory culture (e.g., Schwartz, 2012), and has been highlighted by influential YouTubers like Adam Neely (e.g., 2016) and Jacob Collier (e.g., Lee, 2017). Its prevalence across these sources—as well as in the body of literature analyzed in Part II of this dissertation—suggests that harmonic luminosity might possibly have cultural or cross-cultural properties.

The present experiment aims at examining harmonic luminosity's general perceptual salience. The multitude of theoretical discussions on (and mentions of) harmonic luminosity, and the complete lack of empirical studies, is already a justification of such a study. However, in the context of the present dissertation, the study serves a second purpose: To examine whether harmonic luminosity is a salient cross-modal correspondence for students in higher music education. If so, the concept could represent a concrete teaching method for promoting holistic harmonic listening—especially for learning to aurally recognize and label chromatic mediant

of harmony evoke the *idea* of light and darkness. See Part II of this dissertation for a more thorough discussion of harmonic luminosity's metaphorical nature.

124 "Variant chord" is the equivalent to "parallel chord" in modern anglophone terminology.

125 "Variant, a term introduced in recent years by the author of this lexicon, which is characterized by the substitution of tonic chord quality (major tonic in a minor key, or minor tonic in a major key) by means of alteration of the third. The expression is chosen because such situations do not normally bring about an actual modulation, but rather a sudden brightening or darkening of the key" (my translation).

chords and modal mixture chords, which are often highlighted as particularly expressive of harmonic "brightness" or "darkness." Examples and discussions of such effects are found in Part II of this dissertation.

13.2 Earlier research

The concept of harmonic luminosity is not novel. The idea that harmony can "brighten" or "darken" is found both in academic and non-academic contexts. As shown in Part II, there is also a consensus of sorts about the tonal principles governing these effects. However, while sporadic associations between harmonic movement and brightness/darkness have been found in some studies with a broader scope (see below), I am aware of no systematic explorations of harmonic luminosity's perceptual validity, either in a specialized or in a general population.

This is likely due to the fact that the concept has never been comprehensively and coherently defined. Thus, there has been no established hypothesis to test empirically. The harmonic principles of harmonic luminosity formulated in Part II of this dissertation created a rudimentary framework on which such predictive hypotheses may be built. Accordingly, to my knowledge, the present study is the first of its kind.

Relevant empirical studies can be found in cross-modal correspondences (CMC) research, however. This is a field of studies that has received increasing attention over the last decade. CMC can be described as "systematic associations often found across seemingly unrelated sensory features from different sensory modalities" (Parise, 2016, p. 7), often accompanied by "a phenomenological experience of similarity between such features" (Parise & Spence, 2013). An underlying assumption is that they are shared by a large number of people, or might, in some cases, be universal (Spence, 2011, p. 973). For this reason, among others, CMC are generally treated separately from synesthesia, which has a more idiosyncratic nature. Nonetheless, there is a debate whether synesthesia and CMC are "distinct phenomena, or different points along the same continuum" (Eitan, 2017, p. 219). Two other differences between the phenomena should be mentioned to clarify the concepts: Synesthesia tends to involve an absolute mapping (e.g., specific letters have specific colors), while CMC tend to be relative (e.g., one tone is perceived as vertically *higher than* another, regardless of the specific pitch of either tone)—although see Spence (2019) for a modification of the latter claim. Also, and perhaps more importantly, synesthesia involuntarily *triggers* an unrelated perceptual mode (e.g., D major makes you *experience* yellow) whereas in CMC the cross-modality has a more associative quality (e.g., "lemons are fast, prunes are slow" (Woods et al, 2013)).

Many CMC effects are thought to influence mental processes automatically and subconsciously, without conscious awareness, suggesting they may also "operate under the surface of music processing, implicitly shaping music structures, emotions, and meaning" (Eitan, 2017, p. 216). While it was treated as a *conceptual metaphor* in Part II of this dissertation, it seems reasonable to suggest that if the concept of harmonic luminosity is indeed perceptually valid, it would be a form of cross-modal correspondence between the auditory and visual modalities.

Relevant to the present study, research on CMC has described strong correspondences between auditory pitch and visual brightness. These studies fall into two categories, as summed up by Spence (2019, p. 236). The first is concerned with visual *lightness*, e.g., matching pitches with white, black, or gray stimuli.[126] This group of studies starts with Marks (1974) who demonstrated that all 12 participants in an experiment matched increasing pitch with increasing visual lightness—a cross-correlation he traces back to an observation by Wundt (1874) a hundred years earlier. Several subsequent studies have found the same effect, even including one that demonstrated it in chimpanzees, leading to a proposal that the mapping is a basic feature of the primate sensory system, rather than a culturally learned or linguistic phenomenon (Ludwig et al., 2011). The CMC between pitch and visual lightness has been reaffirmed empirically more recently by Brunel et al. (2015).

The second category mentioned by Spence is concerned with auditory pitch and visual *brightness*, e.g., the amount or quality of light. This is also a well-established correspondence, which Spence traces all the way back to Newton and Goethe, and which is recently described empirically by, e.g., Eitan and Timmers (2010), Klapetek et al. (2012), and Getz and Kubovy (2018). See Spence (2019, p. 236) for a full review.

However, these studies are all concerned with *pitch height*. This basic auditory feature is not specific to music. The difference between pitch and harmony is not only quantitative, but qualitative: Harmony carries both musical information and cultural and emotional connotations, in ways that single pitches do not. Thus, the relevance of research on pitch and brightness to harmonic luminosity is unclear.

At the same time, this is indicative of the status of research on CMC. Eitan points out an important gap in music-related studies within the field:

126 "Lightness" (i.e., light vs. dark *hue* or *color*) in this sense is not identical to "brightness" (i.e., the amount of light). But because of the linguistic and possibly also perceptually overlapping meanings involved (*light/bright* vs. *dark*) these studies may still carry relevance to harmonic luminosity.

> […] even when actual musical stimuli are used (rather than rarified stimuli like single sine tones), CMC research mainly examines music as sound, focusing on basic features shared by most auditory domains, like pitch and loudness. Little CMC research examines intrinsically musical features, such as harmonic and melodic intervals, chord structure, or modality, and CMC involving higher-level musical structures, such as tonality, metric hierarchy, or rhythmic and melodic configuration, are almost completely ignored. (Eitan, 2017, p. 214)

Three studies should still be mentioned in this regard. The first, by Bonetti and Costa (2019), examined classifications of musical mode (major and minor) along five dichotomous scales: happy–sad, pleasant–unpleasant, up–down, light–dark, and warm–cold colors. Participants were 51 children aged 4–6 years, and 168 adults. The major–light, minor–dark association was found to be strong in both groups, with a proportion of 72 percent congruent responses in 6-year-olds and 92 percent in adults (p. 63). While it should be noted that Bonetti and Costa used the parameter *lightness* rather than *brightness* (see above), the robust major–light, minor–dark associations recorded in the study carry obvious relevance to harmonic luminosity, and demonstrate that at least certain types of musical harmony carry strong affordances (cf. Clarke, 2005) for metaphorical descriptions of lightness/darkness.

In the second study, Collier and Hubbard (2004) examined the perceived brightness of different ascending and descending scales. Their findings were that ascending scales were generally rated as brighter than the corresponding descending scales, and keys starting on higher pitches were rated as brighter than keys starting on lower pitches.

A possible weakness in the latter study, however, is the use of the words *dark* and *bright*.[127] These words are clearly polysemic, and it is not clear how or whether Collier and Hubbard were able to ensure that participants really interpreted them in the meaning of "visual brightness." As discussed in Chapter 14.2.1, my own pilot study showed that the labels *dark* and *bright* were significantly contaminated by the overlapping meaning *low* vs. *high pitch*. In other words, participants in the pilot study were considerably influenced by pitch contour when judging the brightness of the harmony,[128] even when they were explicitly told to ignore this aspect. Collier and Hubbard also found that contour influenced ratings of brightness (p. 159), and suggest that pitch contour, along with global pitch height and the pitch distance between subsequent notes, are the main influences for the perceived brightness of musical stimuli (p. 151). Still, it seems reasonable to raise the following question to Collier and Hubbard's study: Was pitch

127 Collier and Hubbard asked participants to rate scales as "dark" or "bright" using a seven-point Likert scale (with "1" as *dark* and "7" as *bright*).

128 This effect was also shown in an experiment by Korsakova-Kreyn and Dowling (2012; 2014) which will be presented shortly.

contour the *cause* of the perceived brightness in the musical stimuli, or could it have been the *consequence* of linguistic confusion over the terms *dark/bright*? In other words, what kind of "brightness" did the participants really rate?[129] Notwithstanding this objection, Collier and Hubbard's experiment is an example of a CMC study involving higher-level musical structures (i.e., modes) and perceived brightness. Their results, though possibly influenced by other associations than visual brightness, are generally also in concordance with expected audiovisual correlations.

In the third study, Korsakova-Kreyn and Dowling (2012; 2014) found that modulations to chromatic step 8 (i.e., modulations from C major to A♭ major) were associated with "dark," while modulations to the dominant were associated with "bright."[130] These findings are in concordance with the luminosity principle presented in the present study. More interesting, however, was the effect of melodic contour: Korsakova-Kreyn and Dowling found that ascending soprano and bass lines contributed to a "bright" effect, while descending soprano and bass lines were associated with "dark." This indicates that pitch contour should be treated consciously in a study on harmonic luminosity. However, the effect was of pitch contour was not identical for all modulations: The *bright/dark* scale for chromatic step 8 was not affected by pitch contour, suggesting the tonal distance was stronger that the influence of the directions of pitch change (Korsakova-Kreyn & Dowling, 2012, p. 554).

In conclusion, the present study aims at simultaneously addressing research gaps in two fields: On the one hand, the need for empirical support for harmonic luminosity, which has been described theoretically for at least 60 years. And on the other hand, the need for more empirical research on cross-modal correspondences involving higher-level musical structures.

13.3 Aims

In an observation that remains relevant, Klonoski (2000) pointed out a discrepancy between the rich and varied research being carried out on music perception, and the practical relevance of this research on aural skills pedagogy:

> To be sure, important cognition research with pedagogical implications has and is being done. Still, the majority of the research that is directly relevant to aural skills

129 One might, for example, have expected a study of this kind to examine differences in perceived brightness between major and minor scales. However, no such (direct) comparisons were made.

130 "Bright" and "dark" were only one of several dichotomous descriptions used in the experiment.

> instruction has focused on isolated perception tasks and understanding both the physiologic and cognitive nature of the processes involved. In short, it has focused primarily on understanding existing cognition skills and processes. Aural skills instruction, on the other hand, explicitly seeks to refine and augment existing auditory processing skills, as well as to develop new ones. (Klonoski, 2000, para. 2)

The present study is an effort to bridge this divide, by providing basic research on a perceptual phenomenon while simultaneously embedding it in the context of aural skills instruction.

The study aims to examine empirically whether musicians and non-musicians experience harmony as expressing "brightness" and "darkness" in accordance with the harmonic luminosity principle put forward in Part II of this thesis. If the principle is found to be perceptually robust, this will serve several purposes: First, it would provide empirical support for a phenomenon which has been described theoretically since at least the early 1960s (see Part II) but whose perceptual validity remains undetermined. Second, it would constitute a widening of the metaphor-oriented basis of holistic harmony identification, from idiosyncratic to more intersubjective metaphors.[131] Third, an intersubjectively valid metaphor for the perceptual experience of harmonic relationships would provide the resources for a concrete approach to holistic aural harmony training. Harmonic luminosity metaphors could be used as exemplary stepping stones in the process of developing holistic identification skills for harmonic material.

To underline the latter aspect, the construction of the experiment strives for ecological validity[132] pertaining to an aural training situation in a classroom setting. This includes the use of instrument (i.e., piano instead of sine tones or Shepard tone generator) and how the topic is communicated to the participants. These issues will be discussed in Chapter 14.

131 The process of developing holistic harmony identification skills through the use of a metaphorical listening approach was proposed and discussed in Part II of this dissertation.

132 "Ecological validity examines, specifically, whether the study findings can be generalized to real-life settings" (Andrade, 2018, p. 498).

14 Method

14.1 Hypotheses and research questions

The hypotheses tested in this study are based on the document analysis of textbooks on harmony performed in Part II of this dissertation. Nine of the 20 textbooks in the sample used metaphorical descriptions of harmonic experience that were in some way or other based on a cross-domain mapping of harmony and brightness/darkness. This cross-domain mapping was termed HARMONIC LUMINOSITY, following the example of Thoresen (2015). All instances of harmonic luminosity metaphors were analyzed, both in terms of metaphorical structure and in terms of harmonic content. A basic principle of harmonic luminosity was extrapolated from this analysis (henceforth referred to as the *harmonic luminosity principle*):

In general, major or augmented intervals above a mode's root create brightness, relative to minor or diminished intervals.

This principle can be conceptualized in two ways. These two conceptualizations also form the hypotheses for the present study:

A. *Adding sharps to (or subtracting flats from) a mode makes it brighter. Adding flats (or subtracting sharps) makes it darker.*
B. *Clockwise modulations along the circle of fifths generally create a brightening effect. Counterclockwise modulations along the circle of fifths generally create a darkening effect.*

While these conceptualizations adhere to the same underlying harmonic principle, they highlight different aspects of harmonic luminosity. Therefore, for experimental clarity, they were treated as separate hypotheses in the experiment. (However, no mutually contradictory results were predicted between the two experimental setups.)

The two hypotheses were the bases of two experimental setups. The first, Session 1, tested the claim about the difference in luminosity between modes, by juxtaposing individual chords that stem from different parent scales. The second, Session 2, tested the claim that luminosity correlates with direction of movement in the circle of fifths, by constructing chord progressions[133] that move strictly in either clockwise or counterclockwise motion.

133 While the claim actually refers to *modulations*, the experimental setup will make use of *progressions*. However, these progressions are relatively slow-moving and consist largely of non-functional chord relations, so the difference

The hypotheses predicted a luminosity effect for each stimulus in the study. If participants' judgements of the stimuli were in accordance with this prediction, it would give support to the hypotheses.

The two first research questions (RQs) were:

1. *Do the two hypotheses consistently predict the participants' judgements of the stimuli?*

2. *How are the participants' performances influenced by musical experience, main instrument, genre, and other factors?*

These RQs addressed the applicability of harmonic luminosity in aural training in higher music education, the primary motivation for including the experiment in this dissertation. However, because empirical studies on harmonic luminosity is virtually non-existent, two additional research questions were formulated. They had the purpose of examining and comparing the luminosity effects of specific types of harmonic movement, thus refining our understanding of the phenomenon. These RQs were:

3. *What types of harmonic movement/juxtaposition cause the greatest luminosity effect, according to participants' judgements?*

4. *In the circle of fifths, which number of positions*[134] *traversed causes the greatest luminosity effect, according to participants' judgements?*

For RQ4, a hypothesis formulated by Mulholland and Hojnacki (2013, p. 167) suggests that the luminosity effect is present already with stepwise movement through the circle of fifths, but successively stronger with leaps of two, three, and four positions respectively. When moving five positions, however, the effect is reversed. This is because five positions along the circle of fifths arrives at a tonic either a chromatic step above or a chromatic step below the starting point. The chromatic movement will itself create a brightening or darkening effect (e.g., moving from C major to C♯ major will constitute a brightening effect due to the sharpening of all pitches) which overrides that of clockwise/counterclockwise movement.

between the two terms is more theoretical than perceptual.

134 One position equaling one "hour" on the clock face, e.g., C major to G major.

14.1.1 Critical remarks regarding the hypotheses

While they do not address the hypotheses directly, two possible counterarguments should be mentioned. The first is a partly rivaling take on the cause and nature of "bright" chords, put forward by music theorist David Huron (2006). The other comes from researcher Charles Spence and concerns the general nature of cross-modal correspondences.

In his book *Sweet Anticipation: Music and the Psychology of Expectation* (2006), Huron presents a collection of qualia for chromatic mediants. These are based on adjectives gathered from "musicians and nonmusician listeners [asked to] provide adjectives that describe their feelings" (p. 271) upon hearing chord progressions involving chromatic mediant chords.

The experiment itself is presented in an anecdotal manner, and Huron offers little if any insight into the methods and stimuli used. For this reason, the qualia he gathers will not be discussed here. However, Huron elaborates his findings into a general hypothesis about the perception of chromatic mediants. This warrants some attention, as it partly opposes the luminosity principle put forward in the present dissertation.

The premise for Huron's hypothesis is the basic observation that chromatic mediant chords create a stronger harmonic contrast than diatonic mediants. This, he argues, highlights certain phenomenological aspects of these chords:

> In light of their poor statistical linkage to the preceding and following chords, the qualities of "major" and "minor" come to the fore. Major chromatic chords tend to sound more distinctly "major," and so are somewhat "brighter" or mor "positive" than major chords within the key. Similarly, minor chromatic chords tend to sound more obviously "minor"—with the consequence that these chords will sound more "serious," "sad," or "tragic" than their diatonic counterparts. (Huron, 2006, p. 274)

This means, for instance, that both ♭III and ♭VI chords in a major key context—considered "dark" chords by the harmonic luminosity principle—are "bright" chords according to Huron's theory. This is apparently also supported by his informal findings.

The harmonic luminosity hypotheses could also be criticized from the perspective of the prevailing paradigm in cross-modal correspondence (CMC) research. Charles Spence (2011) argues that CMCs generally fall into one of three categories: statistical, structural, or semantic/linguistic. He maintains that "[p]airs of sensory dimensions that do not meet any of these conditions … are thus unlikely to exhibit any crossmodal correspondence" (pp. 978–979).

Let us examine these categories with regard to harmonic luminosity. First, a perceptual correspondence between musical harmony and visual brightness is unlikely to be explained by natural statistics of the environment. There are simply no commonplace situations where we are consistently and systematically exposed to certain musical harmonies in conjunction with corresponding light effects. Second, there are no obvious similarities in the structural nature of the two percept categories—that is, commonalities in the neural coding of musical harmony and visual brightness respectively. Spence later grants that very few known CMCs can be explained by the structural account (2019, p. 240). Third, there are no clear semantic similarities in how musical harmony and visual brightness are described through language.[135] Or, rather, there are no such semantic similarities *besides* the phenomenon of harmonic luminosity that could explain how this CMC would arise in the first place. In effect, Spence's tripartite categorization seems to disqualify harmonic luminosity from CMC status.

However, Spence subsequently concedes that there are limitations to this categorization, and that it is "by no means meant to be exhaustive" (2011, p. 989). In an article of 2019, Spence's classifications are no longer presented as *conditions*, but as "classes of explanation" (2019, p. 237). They now also include a fourth category, that of *emotional mediation*. To speculate, the "softer" position in the 2019 article might reflect a general development in CMC research, broadening the scope from simpler to more complex correspondences—correspondences that may require more complex models of explanation.[136]

Finally, harmonic luminosity could conceivably also be a kind of "second-order" association, reflecting a "transitive association" (Spence, 2019, p. 240) between more basic dimensions of perceptual experience.

14.2 Methodological challenges

There are three main challenges to the methodological design of this study. Two of them address the problem of isolating harmonic effects from other aspects of sound that are also

135 Such a semantic overlap does exist between visual brightness and *pitch contour*, as discussed elsewhere in this thesis. However, the principle claims that harmonic luminosity is something fundamentally other than pitches moving in "brighter" or "darker" directions. It should also be mentioned that *color* is a common fundamental metaphor for descriptions of harmony and (certain aspects of) visual light, as thoroughly shown in Part II of this thesis. The color metaphor, however, plays no part in the harmonic luminosity principle, and thus seems irrelevant to the phenomenon.

136 A contribution to the finer categorization of cross-modal experiences and their nature is offered by Macpherson (2011). Macpherson's discussion has a highly philosophical nature, and its applicability to practical research is not obvious to me. However, it further illustrates that Spence's categorizations might indeed not be exhaustive.

associated with the metaphors of "brightness" and "darkness"—specifically those of pitch contour and timbre.

In addition, there is the concern that harmonic luminosity may not be understood intuitively and in the same way by participants in the experiment. On the one hand, measuring it indirectly without revealing the purpose of the study may lead to a misunderstanding of the tasks. On the other hand, revealing too much may be criticized for priming the participants into a specific listening mode. These challenges will be addressed independently in the following.

14.2.1 Harmonic luminosity and pitch contour

A pilot study of 25 harmonic stimuli, performed on 19 fellow PhD students and supervisors, produced interesting results in support of the perceptual reality of harmonic luminosity. However, it raised some concerns about the conceptual clarity of the study. While the majority of stimuli were judged overwhelmingly in accordance with the luminosity principle, a few gained unconclusive results. One stimulus was found by all 19 participants to have the *opposite* luminosity effect as that predicted by the principle.

An examination of all stimuli in this pilot led to a suspicion that the direction of voice leading in the harmonic stimuli—especially in the outer voices—might significantly have influenced the judgement of "brightness" and "darkness." For several stimuli, it was unclear whether the participants were influenced the most by the harmonic relationship of the chords or by ascending/descending pitch contours in the outer voices. This discovery finds support in a study by Korsakova-Kreyn and Dowling (2012; 2014), who also found independent "brightness" effects for tonal movement and pitch contour respectively.

The possible confusion between harmony and pitch contour effects on luminosity is likely due to the ambiguous meaning of the terms *light/bright* and *dark* in the context of music: Besides being metaphors for harmonic luminosity, the terms also—and more commonly—refer to pitch height. The confusion between these meanings may possibly be even greater in Norwegian, the language used in the pilot study, because the adjectives *light* and *dark* (*lys* and *mørk*) are equally common to *high* and *low* as descriptors of pitch height. It is conceivable that overlapping or polysemic terms like these interact in an "abstract semantic network" of cross-modal correspondences, as proposed by Martino and Marks (2001). On the other hand, there is a possibility that confusion between harmonic luminosity and pitch direction (i.e., pitch-verticality) might operate on a purely perceptual level—as a blending of two separate cross-modal correspondences—regardless of native language.

A relevant observation in this regard was made by Eitan et al. (2012): Studies on aural cross-modal correspondences in children suggest that some nonverbal associations appear to come "early and 'naturally,' based on inborn or early learned sensory interconnections, whereas the use of corresponding verbal metaphors matures later, and may initially even hinder mappings previously acquired nonverbally." (p. 34) Even in an adult population, it should not be ruled out completely that linguistic categories—like the polysemic terms "light/dark"—may actually confound the judgement of harmonic luminosity.

While the effect of pitch contour was most apparent in the group of stimuli that gained unconclusive results or even contradicted the luminosity principle, it also raised concerns about the validity of the conclusive results. Consciously or unconsciously, the participants in the pilot study may have been influenced by pitch contour even in the cases where they judged in accordance with the luminosity principle: If a stimulus contained *both* a harmonic darkening *and* a descending voice leading pattern, it was difficult to assess the perceptual strength of the harmonic darkening independently from the voice leading.

The discovery of pitch contour effects in the pilot study led to several methodological revisions in the final experiment. Some of these concerned the construction of stimuli and led to measures being taken to minimize the perceptual effect of voice leading. These measures are described in detail in Chapter 14.3.

14.2.2 Timbre and spectral centroid

Drouzas and Saitis (2020) note that words related to the notion of brightness (such as *bright*, *dark*, *dull*, *brilliant*, or *shining*) are among the most common descriptions of timbre used by musicians, music engineers, audio researchers and everyday listeners. This observation is also made by Walther-Hansen (2020, p. 98). Therefore, a study of the perception of harmonic brightness should consider what role timbre may play in the experimental setup. In the present study, the impact of timbre is reduced by adhering to one instrument throughout the stimuli. Instrumental timbre, however, is only one facet of this methodological challenge.

The *spectral centroid* is a measurement which indicates where the center mass of a sound's spectrum is. It is calculated as the weighted mean of the frequency spectrum. The spectral centroid is a good predictor of a sound's perceived brightness (Grey & Gordon, 1978; Schubert & Wolfe, 2006)—the higher the centroid value, the "brighter" the sound is generally perceived. The "brightness" associated with a higher spectral centroid value, however, is not the same kind of brightness as that which is postulated in harmonic luminosity. In textbooks on harmony, harmonic luminosity is portrayed as an effect stemming from *tone content*—a sense

of brightening or darkening due to modal or chordal contrast. The brightening associated with spectral centroid, on the other hand, it is a *timbral* effect which stems from a greater emphasis on higher frequencies in the sound's spectrum.

Still, it is difficult to separate these two kinds of brightness completely from each other. While a high-pitched sound is not necessarily brighter than a low-pitched sound (Walther-Hansen, 2020, p. 99)—this depends on the frequency contents of the respective sounds—a higher pitched note generally has a somewhat higher spectral centroid than a lower pitched note played *on the same instrument* (Schubert et al., 2004). Consequently, any chord change is likely to involve some degree of shift in the placement of the spectral centroid, simply because the spectrum itself is not identical between the two chords. This illustrates the conceptual and auditory entanglement of "brightness" effects stemming from harmony, pitch movement, and timbre/spectrum respectively.

In conclusion, perceivable shifts in the spectral centroids of the stimuli could possibly impact the results in the study. There is a possible brightness effect in any chord change that is independent from harmonic luminosity, and which cannot be eliminated altogether without eliminating pitch changes (and thereby, the very possibility of chord changes in the first place). In any given case, these two effects will either work in the same direction (e.g., harmonic brightening *and* spectral brightening) or they will pull in opposite directions. A methodological concern, regarding construct validity, was to avoid undue reinforcement of luminosity effects by spectral effects in the stimuli.

Several measures were taken to counter this, and to generally minimize dramatic spectral centroid variation. In addition to the use of only one instrument (digital piano) with minimal voice movement, these include the use of a limited, uniform register, uniformity of note onset and velocity for every chord in every stimulus, and consciously making noticeable shifts in spectral centroid placement "move" in the *opposite* direction of the predicted harmonic luminosity, wherever such a shift was unavoidable. However, no technical calculations of spectral centroids were made.

14.2.3 Conceptual clarity

In the pilot study, the participants were told openly that they were to judge the harmonic brightness/darkness in the listening tasks, along with a short description of the phenomenon. As it turned out, it was still challenging to ignore completely the effect of voice leading (see above). However, one might argue that revealing the concept of harmonic luminosity *prior*

to the experiment primed the participants to performing better (that is, more in accordance with the hypothesis) than if they had not received such instructions.

This was a dilemma that also had to be handled in the final study. On the one hand, priming the participants to a special listening focus could be criticized for reasons of validity. On the other hand, not providing participants with sufficiently clear instructions carries the risk of them misunderstanding the listening tasks. In that case, it would be difficult to know how they interpreted the tasks, and thus, what perceived question they had really answered.

A possible solution that was contemplated was to exclude linguistic descriptions of "brightness" and "darkness" altogether, and instead communicate these qualities through corresponding visual symbols (e.g., images of a sun and a dark cloud respectively). Matching visual symbols with auditory stimuli might be an intuitive task that would not demand an introduction the concept of harmonic luminosity. However, this solution did not seem to convincingly address the possible confusion between luminosity and pitch contour effects which was an issue in the pilot study. It seemed necessary to guide participants' attention towards harmonic effects rather than pitch effects.

In the final study, a combination of approaches was used: The concept of harmonic luminosity was briefly introduced[137] before the listening tasks, to make participants (somewhat) attuned to the phenomenon. In addition, the participants expressed their judgement of the stimuli using image buttons rather than the words "brighter" and "darker." These measures were intended as a compromise between guided and intuitive listening. This decision is justified from the perspective of ecological validity: Since one of the aims of the study is to examine whether harmonic luminosity is a relevant concept for aural training classes (i.e., are music students able to perceive harmonic brightness and darkness according to the hypotheses), the concept should be introduced the same way in the experiment as it would in an aural training setting—through a combination of perceptual guidance and perceptual intuition.

14.2.4 Accusations of psychoacoustical reductionism

In addition to the aforementioned challenges concerning its methodological design, some might consider problematic the study's treatment of harmonic luminosity as something that can be studied in isolation. By excluding or disregarding other relevant aspects like timbre, pitch, instrumentation, chordal overtone spectrum, tuning, musical context, performance, etc., the experimental design may be vulnerable to accusations of psychoacoustical reductionism.

137 No theoretical information about the concept was given, only a short description of the association to "light" and "darkness" in the perception of harmony.

This is a valid concern. While formal research is lacking, both Lasse Thoresen's[138] and my own anecdotal studies of harmonic luminosity in Western music indicate that such harmonically driven effects are, indeed, often reinforced by other musical parameters. For example, a distinct harmonic brightening may be accompanied by shifts in pitch register and orchestral timbre—all three of which, independently, communicate a sense of psychoacoustical "brightening." On the one hand it is possible, due to the conceptual overlaps of these and yet other psychoacoustical properties, that the discernment of harmonic luminosity, as it is presented in the literature, may require some degree of perceptual learning. In other words, one may have to learn to discriminate between different "kinds" of auditory brightness before one can perceptually isolate and consistently recognize those stemming exclusively from harmonic relations. On the other hand, it is also conceivable that what the literature portrays as a purely harmonic phenomenon *is* in fact a composite phenomenon, in which voice leading and timbre also play some part.

Nonetheless, as I have shown in Part II of this dissertation, harmonic luminosity as a conscious concept can at least be traced back to Persichetti's 1961 book. Regardless of concerns of psychoacoustical reductionism, the fact remains that in the 60 years following its publication, the concept has been echoed, adapted and developed by numerous sources and in several musical contexts. In all sources I am aware of—those analyzed in Part II as well as other presentations of the concept—harmonic luminosity is clearly presented as an *autonomous* effect, unaffected by influences outside the realm of harmonic relationships. The fact that it is presented as autonomous is also, implicitly, the main justification of harmonic luminosity's theoretical and pedagogical value.[139]

For this reason, the present experiment treats harmonic luminosity as independent from other (and possibly conceptually overlapping) parameters.[140] If, instead, harmonic luminosity is indeed shown to be contingent on specific voicings or pitch contours, on the timbre of the performing instrument, etc., then this is an important clarification of the concept.

14.3 Stimuli

All stimuli were made using Cubase 10.5 software with a naturalistic piano plugin (Alicia's Keys) using equal temperament. All chords were performed as whole notes at 120 bpm. All

138 In private conversations.

139 Although as a creative concept for composers, harmonic luminosity is less contingent on such autonomy.

140 Although, as discussed, the influence of other parameters is minimized in the experimental design.

three-chord stimuli (33 of 35) lasted 6.22 seconds, while five-chord stimuli (2 of 35) lasted 10.17 seconds. Note onsets and velocity were uniform for all chord tones and across all stimuli.

The stimuli predominantly consisted of four-part chords. The register used was predominantly octaves 3–4. The register was allowed to vary somewhat between stimuli, but to a significantly lesser degree within stimuli. Db2 (MIDI number 37) and D5 (MIDI number 74) were the lowest and highest tones respectively across the stimuli. A notated overview of all stimuli can be found in Appendix 2.

All stimuli were forced-choice tasks, requiring the participants to make a binary judgement about the perceived brightness or darkness, e.g., "in your experience, does this chord progression move from *dark to bright*, or from *bright to dark*?" Participants made their choice by clicking one of the buttons that represented these options. One option always represented the luminosity effect predicted by the hypothesis, while the other option contradicted the hypothesis. Simplicity and clarity were favored over nuance due to the methodological challenges already discussed, and participants were not given the chance to grade or quantify their luminosity experience of each stimulus. However, they had the opportunity to answer "Sounds about the same" (Session 1) or "Vague / No clear change either way" (Session 2) if they found the luminosity effect to be weak, unclear, or ambiguous, or "I don't know" if they had no opinion. The experimental setup can be viewed through Gorilla Open Materials.[141]

The possible influence of voice leading on the perception of harmonic luminosity has already been discussed. In the final study, this was addressed in several ways. Generally, in studies on the perception of harmony, an established way of obscuring voice leading effects is to use Shepard tones. Shepard tones are "composed of sine-wave, octave-spaced components over a seven-octave range using an amplitude envelope that tapers off at both low and high ends of the frequency range," which "greatly reduces the clarity of melodic gestures, voicing, lowest and highest pitch, and chord inversions" (Jimenez & Kuusi, 2018a). For these very reasons, however, music performed on a Shepard tone generator takes on a very special tone color.

While a Shepard tone generator can be likened to a digital organ in terms of sound, both the timbre and the obscuring of octaves may appear foreign to unaccustomed ears. For this reason, the idea of using Shepard tones was ultimately discarded. The main argument for the decision was that of ecological validity: The study aims at examining the potential for using harmonic luminosity as an aural training concept in the classroom. Using Shepard tones would potentially remove the stimuli so far from an everyday musical context—and from the physical context of the classroom—that it could undermine one of the purposes of the study.

141 https://app.gorilla.sc/openmaterials/340587.

Rather, the methodological challenge of voice leading has been met through careful construction of the stimuli. In the majority of stimuli, the pitch contour of the top voice contradicts the predicted luminosity (see discussion in greater detail below). This does not eliminate the possible challenge of voice leading entirely, but it reduces the risk of false positive results.

14.3.1 Session 1: ABA patterns

Session 1 tested hypothesis A, and consisted of 16 three-chord stimuli in ABA pattern. In these stimuli, the *A* chords were intended to act as a tonally "neutral" element, which should not in itself possess any distinct luminous quality either way. To achieve this neutrality—and to minimize the effect of harmonic "coloring" of *A* by harmonic movement from stimulus to stimulus—the *A* chord across all stimuli was placed in the key of G or G minor.[142] While simple ABA patterns are not enough to create a clearly defined tonal context (as would, for instance, a perfect authentic cadence), the relative stability of *A* chords across the stimuli was intended to strengthen G as a perceived tonic. The two *A* chords in each stimulus were always identical to each other in terms of structure and voicing, although they varied somewhat between stimuli.

The *B* chord was intended to create a tonal contrast against the neutral backdrop of *A*, deriving harmonic brightness or darkness from the juxtaposition. In most of the stimuli, the *B* chord was either a modal mixture chord or a chromatic mediant chord. In other words, *B* chords stemmed from a different parent scale than *A* chords. A few of the stimuli used a simpler schema whereby a single chord tone in *A* was sharpened or flattened chromatically in *B*. All *B* chords were musically conventional in Western music,[143] although several were more typical of classical Western music than of the harmony found in pop/rock.

In Session 1, the top voice was generally assigned common tones in 13 of 16 stimuli. In the remaining three stimuli, the top voice was made to move in the *opposite* direction of the predicted luminosity (i.e., *up* when the predicted luminosity was "dark", and *down* when the predicted luminosity was "bright"). For the middle voices, stepwise voice leading was preferred, but movement was allowed in either direction regardless of predicted luminosity. For the bass, larger leaps were permitted. Like with the top voice, bass movement was generally

142 There is naturally a risk that the transition from a stimulus in a major key to a stimulus in the parallel minor key (or vice versa) could *itself* cause a global luminosity effect for the latter stimulus. However, as stimuli in both major and minor tonality contexts were necessary for the study, this option was judged to be the least intrusive. In addition, randomization of the sequence of stimuli for each participant made such conceivable effects unlikely to systematically influence the result.

143 That is, a chord that could naturally follow *A*, although in a musical situation it might normally be followed by a chord *C*, rather than returning to *A*.

made in the opposite direction of the predicted luminosity, except in some instances where this would create a distinct shift in register (e.g., a descending sixth instead of an ascending third).

14.3.2 Session 2: Graded progressions

Session 2 tested hypothesis B. It consisted of 19 stimuli, of which 17 were three-chord progressions and two were five-chord progressions. Every progression was based strictly on either clockwise or counterclockwise movement through the circle of fifths.[144] Hypothesis B predicted that the former would induce a sensation of harmonic brightening in the listeners, while the latter would induce a sensation of harmonic darkening. The aim of Session 2 was to test this claim.

The graded progressions were constructed primarily using triads in root position. Inversions and added tones were used in a minority of the stimuli. Unlike in Session 1, the stimuli in Session 2 did not start and end with the same chord. Neither did any stimulus presuppose the perceptual presence of a "neutral" or tonic chord. Therefore, no effort was made to establish any sense of global tonality across the stimuli. On the contrary, the tonal starting points of the progressions were varied intentionally, to encourage a sense of "fresh start" for each stimulus. The sequence of stimuli was then randomized for every participant.

A secondary aim was to examine whether the luminosity effect varied with the distance traversed in the circle of fifths. Mulholland and Hojnacki propose that the effect is successively greater in movements of one, two, three, and four positions, but is reversed when leaping five positions (2013, p. 167). To test this, 10 of the progressions were made with a systematic approach, with constant structure chords (e.g., only major chords in root position) moving by either one-, two-, three- or four-position patterns (e.g., major triads moving counterclockwise by minor thirds). The remaining nine progressions were constructed more freely and were informed by harmonic conventions of Western classical music. They used a mix of chord qualities and structures, moving by unequal number of positions through the circle of fifths—but still strictly clockwise or counterclockwise.

For the two progressions in Session 2 consisting of five chords, the top voice was constructed the same was as in Session 1, i.e., with movement in the opposite direction of the predicted luminosity. In the three-chord progressions, a more natural melodic contour was allowed,

144 Hence the term *graded progressions*: Rather than the pendulum movement used in the ABA patterns of Session 1, Session 2 consisted of ABC(DE) patterns where each stimulus was a *gradually* brightening or darkening chord progression respectively.

and smooth, stepwise voice leading was favored over the systematic contrasting of predicted luminosity.

14.3.3 Recruitment of participants

Statistical research is often based on probability sampling. De Vaus (2014) describes this as a sampling strategy in which neither the researcher nor the participants have any influence on who invited to participate. Probability sampling is necessary if the study aims to achieve results that can be generalized to the wider population. However, this aim is not relevant or convenient to all studies, for example exploratory studies, or studies where a certain category of cases are considered more interesting than others (De Vaus, 2014, p. 88).

Because of the present study's exploratory nature and the explicit context of aural training, a non-probability sample was chosen. However, the sampling strategy followed some principles: People who were presently, or had formerly been, students in higher music education (HME) were identified as a group of special interest. They were judged to better represent aural training students in higher education than would a random sample, and the perceptual salience of harmonic luminosity in this group was thus considered more relevant to the study's aims. This is described by De Vaus as *purposive* sampling (De Vaus, 2014, p. 88).

However, HME also served a second purpose in the study. It is possible that the judgement of harmonic luminosity may require some degree of perceptual learning. In such an eventuality, HME students represent a part of the general population that is more likely to be attuned to the perceptual nuances in question (i.e., they are expected to possess a higher degree of musical sophistication than the average population). Including a substantial group of music students in the study could thus reduce the chances of a falsification of the hypotheses resulting from targeting a too diverse sample. That is to say, the harmonic luminosity principle may show perceptually valid for a subpopulation of musically sophisticated participants, even if it does not for a general population sample. HME students were chosen as a way of operationalizing such musical sophistication. Formal tools for this exist, like the Goldsmiths Musical Sophistication Index[145] (Gold-MSI). However, the Gold-MSI specifically aims at assessing musical sophistication in a *non*-musician population (Müllensiefen et al., 2014), and would furthermore significantly lengthen the time needed to complete the survey. Instead, musical education itself was used as a proxy variable in this study. Current or former participation in an HME program was determined an acceptable threshold measure for an above-average level of musical sophistication.

145 www.gold.ac.uk/music-mind-brain/gold-msi/

A goal of 200 participants was set, based on considerations of feasibility. Of these, a quota of at least 50 percent was reserved for current or former students in HME, with the remainder of participants having no higher education in music.

The experiment was created as a web-based survey, using Gorilla Experiment Builder[146] (Anwyl-Irvine et al., 2020). This strategy was chosen for several reasons: First, it was an adaptation to the lockdown and social distancing requirements brought about by the Covid-19 pandemic. Second, it eased recruitment of participants from countries other than Norway. And third, it was able to target students in HME specifically.

An invitation to participate in the experiment through a simple link was distributed to a range of HME institutions. These included all public HME institutions in the Nordic countries, as well as several others in Europe and USA. Once about 100 participants had been recruited using this method—ensuring that a substantial part of the total sample was likely to be music students—the link was shared freely through other channels, particularly Facebook. Thus, the final sample also contained a substantial group of participants who were *not* students in HME, allowing for analyses of subsample differences.

Participation was not compensated, and the experiment was based on voluntary response. While self-selection samples like these can be useful for exploratory research, it must be kept in mind that they are unlikely to produce representative results (De Vaus, 2014, p. 89). The opportunity to participate in the study lasted from September 11 through October 7, 2021.

14.4 Ethical considerations

The lower age limit for participation in the experiment was set to 18 years. Complete anonymity was guaranteed through using Gorilla Experiment Builder, which is built to support the existing standards of The British Psychological Society[147] and National Institute for Health Research[148], as well as the legislative requirements of the GDPR[149] and the Data Protection Act 2018[150]. Access to the experiment was distributed through a generic link and did not

146 The Gorilla Experiment Builder (www.gorilla.sc) is "an online experiment builder whose aim is to lower the barrier to access, enabling all researchers and students to run online experiments (regardless of programming and networking knowledge)" (Anwyl-Irvine et al., 2020, section 7).

147 www.bps.org.uk/

148 www.nihr.ac.uk/

149 gdpr-info.eu/

150 www.legislation.gov.uk/ukpga/2018/12/contents/enacted

require any registration or use of participants' e-mail addresses. Since no geographical limitations were used in the experiment, participants' IP addresses were neither registered nor stored by Gorilla. These conditions satisfied the standards for anonymous data processing formulated by the Norwegian center for research data (NSD), and thus the experiment was not subject to notification.

All participants were required to sign a digital consent form before accessing the experiment (see Appendix 1).

14.5 Participants

236 participants completed both Sessions 1 and 2 of the survey and were included in the study. Another 242 persons were rejected because they did not complete the survey, and 8 were rejected because they were under the age of 18.

The 236 included participants (141 male, 91 female, 1 transgender/non-binary/other, 3 preferred not to answer) had an age range of 18–70 years, with a mean of 34.80 years. For participants who reported that they were current or former students in HME (n = 148), information about their institution's nationality was queried. Seventeen different countries of education were reported, the five most common being Norway (n = 71), Denmark (n = 22), Sweden (n = 16), Germany (n = 10) and the United States (n = 9).

	Music education	Hobby / amateur musicians	Non-musicians	Total
Female	56	17	18	91
Male	88	23	30	141
Prefer not to answer	3	0	0	3
Transgender / non-binary / other	1	0	0	1
Total	**148**	**40**	**48**	**236**

Figure 12: Overview of participants

15 Results

15.1 Choice of analysis strategy

The analysis concentrated on measuring the percentage of hypothesis-supporting answers. For each stimulus, the answer option predicted by the hypothesis was assigned the value *1*, while all three other answer options were assigned the value *0*.[151] This method has the advantage that it allows for a simpler statistical analysis which addresses the research questions in a straightforward manner. Hypothesis-supporting answers were simply added together and treated as a performance score (16 for Session 1, and 19 for Session 2, totaling a maximum of 35 "points").[152] This allowed for measuring mean performances and made it feasible to compare the performance of different subgroups to each other. At the same time, it also showed clearly which stimuli achieved the highest and lowest concurrence with the hypothesis.

This strategy does have a drawback, however. For stimuli where a minority of participants chose the hypothesis-supporting answer option, this method does not distinguish between the answer options chosen by the majority. This means that the analysis does not take into account whether these participants chose the hypothesis-contradicting option, the "I don't know" option, or the "Vague / No clear change either way" option. The "I don't know" option was generally chosen relatively infrequently in the survey (2.2% of replies in Session 1 and 1.14% of replies in Session 2). The "Vague / No clear change either way" option, however, was chosen 12.95% of the time in Session 1 and 25.22% of the time in Session 2. Therefore, it is possible that incorporating the "Vague / No clear change either way" option in the analysis would have uncovered interesting nuances in the results, especially in the graded progressions stimulus set. The information is obtainable from the data set and could be analyzed in a future article with a stricter focus on the perceptual nature of harmonic luminosity. For the context of the present study, which is to evaluate the relevance of harmonic luminosity as a methodological tool for aural training (cf. RQs 1–2), these nuances are of less importance. For this purpose, the data is more easily assessed with the described statistical methods.

The complete data set is available on request by contacting villelangfeldt[at]gmail.com.

151 The other options were, respectively: the hypothesis-contradicting option, "I don't know," and "Vague / No clear change either way."

152 In the following discussion, a "higher score" is thus synonymous with being more in accordance with the hypothesis.

15.2 Descriptive statistics

The performance on Session 1 and Session 2 respectively, and the two sessions combined, were analyzed by adding together the number of hypothesis-supporting answers.

For Session 1 (ABA patterns), the mean performance of all participants was 8.33 hypothesis-supporting answers. For Session 2, the mean performance was 8.83 hypothesis-supporting answers (see Figure 13).

	N	Min	Max	Max. possible	Mean	Std. dev.
Age	230	18	70		34,80	12,41
Session 1 total score	236	1	15	16	8,33	2,34
Session 2 total score	236	1	14	19	8,83	2,40
Combined score	236	4	27	35	17,17	3,96

Figure 13: Descriptive statistics of total scores

While the mean performance is higher in Session 2 in sheer numbers, one must bear in mind that the latter session included more tasks than the first (19 vs. 16 tasks). Calculating the mean performance by percentage instead shows that Session 1 had 52.1 percent hypothesis-supporting answers, while Session 2 had 46.5 percent. The data was subjected to a test of normality, using the Shapiro-Wilk test. This test showed that the data from both Session 1 ($W = 0.977$, $p = 0.001$) and Session 2 ($W = 0.976$, $p = 0.000$) deviated from a normal distribution ($p < 0.05$). Consequently, t-test could not be used for comparing means. Wilcoxon signed-rank test, a non-parametric alternative, was used instead. This test also suggested that the scores on Session 1 and Session 2 differed significantly ($Z = -3.043$[153], $p = 0.002$). From this, we can conclude that participants generally scored higher (that is, answered more in accordance with the hypothesis) on the ABA tasks in Session 1 than on the graded progressions in Session 2.

153 Based on positive ranks.

15.3 Combined total performance mean separated by gender

An analysis of combined total performance mean separated by gender was conducted, in order to rule out any gender differences.

The Kruskal-Wallis H test is a rank-based nonparametric test that can be used to determine if there are statistically significant differences between two or more groups of an independent variable on a continuous or ordinal dependent variable. It is considered the nonparametric alternative to the one-way ANOVA, and an extension of the Mann-Whitney U test to allow the comparison of more than two independent groups.

A Kruskal-Wallis H test showed that there was no statistically significant difference in combined total performance between genders.

15.4 Combined total performance mean separated by musical background

Participants were divided into three groups based on musical background. The first group, "educated musicians" (n = 148; 88 male, 56 female, 1 transgender/non-binary/other, 3 preferred not to answer) consisted of participants who were currently (n = 100) or had formerly been (n = 48) a student in HME (university, conservatory, college, or similar). The second group, "hobby/amateur musicians" (n = 40; 23 male, 17 female), consisted of participants who had not studied music at a higher education level, but still considered themselves musicians at some level.[154] The third group, "non-musicians" (n = 48; 30 male, 18 female), consisted of participants who did not report belonging to either of these categories.

The results from Session 1 and Session 2 were combined into a total performance score (maximum score 35). Separating the participants by musical background (see Figure 14), we see that educated musicians had a mean of 17.95, with a standard deviation of 3.76. Hobby/amateur musicians had a comparable score (M=17.77, SD=3.05), while non-musicians scored lower (M=14.25, SD=3.96). The total mean across all participant groups was 17.17, with a median of 17 and a standard deviation of 3.96.

154 Regrettably, this category did not take into account that people without higher music education may still have a professional music career. Based on feedback in the comment section of the experiment, it should thus be noted that the "hobby/amateur" category does in fact contain (at the very least) a small number of professional musicians.

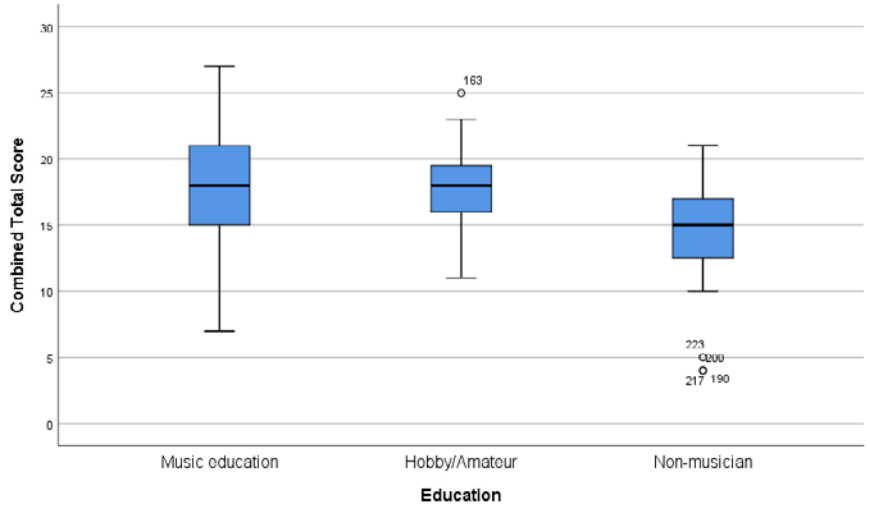


Figure 14: Mean separated by musical background

A couple of initial observations can be made. Educated musicians, though having a higher mean score than the two other groups, is an uneven group. The hobby/amateur group performs somewhat more consistently, and at almost the same level as educated musicians. The non-musician group scores markedly lower than the two groups of musicians. At first sight, these numbers seem to indicate that musical training in some form or other has a positive effect on the judgement of harmonic luminosity. However, even the educated musicians' mean of 17.95 constitutes only 51.29% hypothesis-supporting answers, so the strength of these results should be viewed with some skepticism.

Secondary analyses were performed to determine whether the differences in total performance mean between groups were statistically significant. These analyses showed that the results of educated musicians and hobby/amateur musicians did not statistically differ (non-significant). However, educated musicians' and non-musicians' performances did statistically differ ($t_{194} = -5.84$, $p < .001$), as did those of hobby/amateur musicians vs. non-musicians ($t_{86} = -4.604$, $p < .001$).

15.5 Effect of other factors on combined total performance

15.5.1 Absolute pitch

While possessors of absolute pitch (AP) were regarded as equally suitable participants in the experiment, such listeners are prone to becoming outliers in any study on musical perception and should thus be identified.

However, operationalizing AP poses some methodological challenges. While the term itself indicates that AP is, indeed, an *absolute* capability, it is scientifically regarded more as a skill along a certain spectrum, and might even be contextual (Hedger et al., 2015). Studies have also found it to be influenced by timbre (e.g., Marvin & Brinkman, 2000; Miyazaki, 1989). However, this is not necessarily common knowledge, even among musicians. Consequently, the question "Do you possess absolute pitch?" might not be easily answerable by either "yes" or "no." In order to treat AP as a dichotomous variable in this study, the question used in the survey was softened somewhat, to "Do you experience having absolute pitch?", slightly shifting the power of definition over to the respondent. In addition, the respondents were given the possibility to answer either "Difficult to answer (please specify)" or "I'm not sure what absolute pitch is" in addition to "yes" or "no." Also, the background question about AP was presented only to participants who belonged to the educated musician group, as well as to hobby/amateur musicians who had received formal music training (see 15.5.4). This was done to avoid confusion over the term by participants who might not be familiar with the term.

10 participants reported having AP. 10 additional participants chose the "Difficult to answer (please specify)" reply option, but after closer inspection of their written specifications, these participants were not included in the AP subsample.

No statistically significant correlation was found between AP and performance.

15.5.2 Musical genre of higher music education program

The educated musicians were asked about the musical genre of the HME program they attended (Figure 15). An expectation was that participants with a background from jazz music might gain a higher score than classical musicians, due to the relatively stronger focus on both harmonic construction and awareness in many jazz styles. However, no statistically significant correlation was found between musical genre and performance.

	Frequency	Percent
Folk / etnic	1	0,7
Church music	1	0,7
Jazz / improvised music	28	18,9
Multiple	10	6,8
Pop / rock	7	4,7
Western classical music (incl. contemporary classical)	101	68,2
Total	**148**	**100,0**

Figure 15: Musical genre of HME programs

1.1.1 Main instrument

The group of educated musicians were asked about their main instrument (Figure 16). It was expected that musicians with a keyboard instrument as their main instrument might have a higher degree of harmonic awareness than others, and could thus be better attuned to harmonic luminosity. However, no statistically significant correlation was found between main instrument and performance.

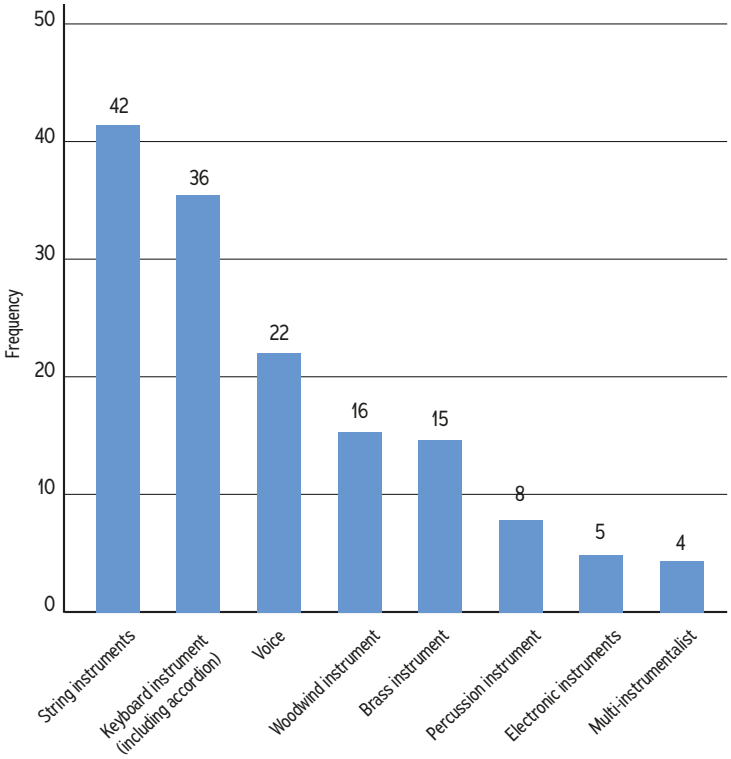


Figure 16: Main instrument of educated musicians

1.1.2 Formal music training

Hobby/amateur musicians were asked whether they had ever received formal music training (e.g., private lessons, evening classis, academic courses, etc.), and for how long (Figure 17). No statistically significant correlation was found between formal music training in the hobby/amateur musician group and performance.

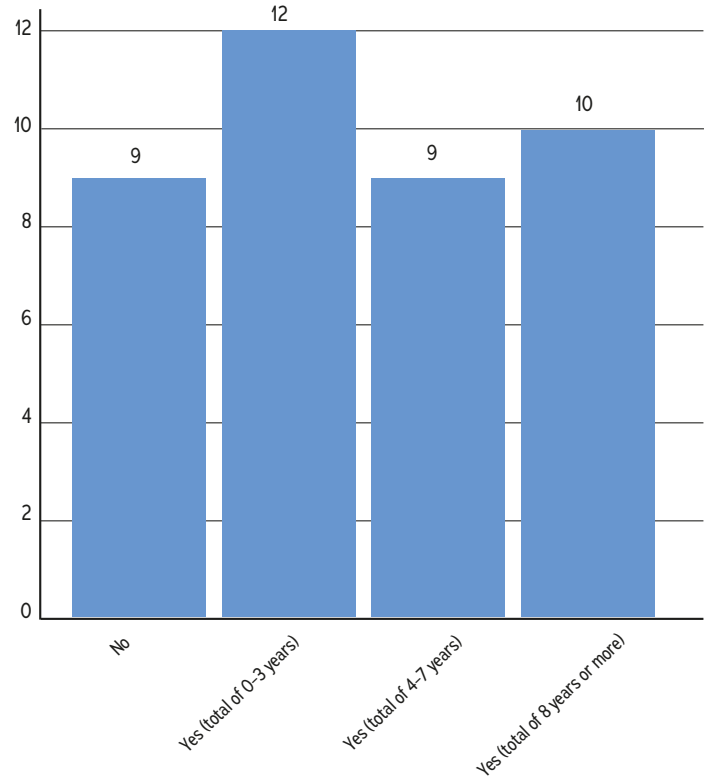


Figure 17: Formal music training among hobby/amateur musicians

15.6 Analyses of themed stimulus sets

15.6.1 Themed stimulus sets

After the initial analyses of Session 1 and Session 2 separately, all individual stimuli were organized into sets according to type of harmonic relation. A set could contain stimuli from both sessions. The purpose of this was to examine the luminosity effect of specific harmonic relations, in order to better answer RQ3 and RQ4. The following stimuli sets were constructed:

1. **Chromatic mediants:** 8 stimuli containing chromatic mediant relations.

2. **Modal mixture:** 5 stimuli indicating either modal mixture, Dorian, or Mixolydian.

3. **Chord tone change:** 3 stimuli containing chords were only one tone changed at a time (e.g., from major to minor).

4. **Circle of fifths, one position (COF1):** 2 stimuli with progressions moving stepwise through the circle of fifths.

5. **Circle of fifths, two positions (COF2):** 4 stimuli with progressions moving through the circle of fifths by leaps of two positions.

6. **Circle of fifths, three positions (COF3):** 4 stimuli with progressions moving through the circle of fifths by leaps of three positions.

7. **Circle of fifths, four positions (COF4):** 4 stimuli with progressions moving through the circle of fifths by leaps of four positions.

8. **Other/mixed:** 5 stimuli which did not fit into any of the former categories, or which combine several harmonic relations.

15.6.2 Stimulus set group mean scores across educated and amateur musicians

Analyses of the themed stimulus sets show some variation in performance between groups, but the variation appeared to be rather random. Hobby/amateur musicians had a slightly higher mean of hypothesis-supporting answers than educated musicians on *chromatic mediants* (58.13% vs. 54.05%) and *modal mixture* (46.5% vs. 42.03%), but educated musicians scored markedly higher than hobby/amateur musicians at *chord tone changes* (81.31% vs. 60.83%). In COF progressions, hobby/amateur musicians scored higher than educated musicians in all stimulus sets except *COF2*, where educated musicians had a mean of 52.2% vs. hobby/amateur musicians' 40.63%. For the *Other/mixed* stimulus set, educated musicians and hobby/amateur musicians scored consistently high, with 77.97% and 76.5% hypothesis-supporting answers respectively.

In conclusion, the group differences in performance between the educated and hobby/amateur musicians did not seem to follow any particular pattern or tendency. The results were not uniform enough to suggest any consistent differences between the two groups, although they both scored higher than the non-musician group.

The specific harmonic contents of the stimuli in each themed stimulus set were then analyzed, in order to examine whether the data suggested answers to RQ3: *What types of harmonic movement/juxtaposition cause the greatest luminosity effect, according to participants' judgements?*

15.6.3 Chromatic mediants stimulus set

For the *Chromatic mediants* stimulus set, the mean across all participant groups was 4.39 hypothesis-supporting answers of a possible 8 (54.87%). This was the third strongest result for any stimulus set. The strongest stimulus in the set, by a large margin, was Grade1 (Figure 18). This progression was rated as *brightening*—in accordance with the hypothesis—by 89 percent of all participants (educated musicians 90.5%, hobby/amateur musicians 85%, non-musicians 87.5%). However, this score was so much higher than other stimuli in the set that it raised the question of whether the ascending pitch contour of the top voice may have reinforced the results (see the discussion on pitch contour in Chapter 14.2.1).

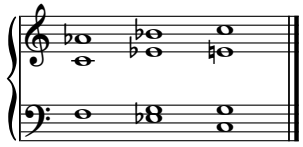


Figure 18: Grade1

The second strongest stimulus in the set was ABA4 (Figure 19), which scored 77.0/75.0/68.8 in the three groups. Unlike the previous example, this result was not likely influenced by pitch contour. Thus, it is possible that the ♮VI chromatic mediant is generally perceived as a brighter harmonic function than other chromatic mediants. The two stimuli using the ♮III mediant[155] (a mean of 38.1% and 64.4% respectively across all participants), and the three stimuli using ♭VI chromatic mediants (a mean of 44.9%, 23.3%, and 69.9% respectively across all participants), showed internally inconsistent results, indicating that these harmonic relations had no consistent luminosity effect in any participant group.

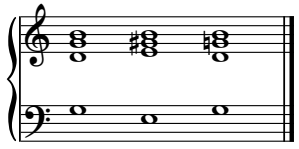


Figure 19: ABA4

155 The last chord in Grade1 (see above) may also be interpreted as a ♮III mediant in A♭ major, but I have not included it in this group due to its tonal ambiguity.

15.6.4 Modal mixture

For the *Modal mixture* stimulus set, the mean across all participant groups was 2.04 of 5 possible hypothesis-supporting answers (40.76%). ABA2 (Figure 20) was relatively strong in both musician groups, but weak in non-musicians (educated musicians 67.6%, hobby/amateur musicians 67.5%, non-musicians 22.9%). ABA15 (Figure 21), which contained a minor subdominant contrasted with a major tonic, was strong in all groups (educated musicians 73.6%, hobby/amateur musicians 85.0%, non-musicians 64.6%)—but ABA14, a variant of the same, was not. The latter case might be caused by contrasting, ascending pitch contour in both outer voices, which could override the effect of the harmonic relation.

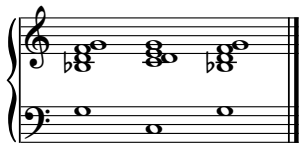


Figure 20: ABA2

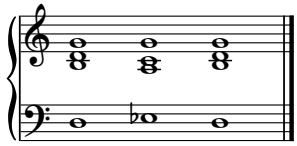


Figure 21: ABA15

Interestingly, ABA16 (Figure 22) was very weak in all groups (educated musicians 15.5%, hobby/amateur musicians 27.5%, non-musicians 20.8%), suggesting the middle chord was perceived by participants as darkening, in spite of introducing a brighter modality, according to the hypothesis. This could indicate that the "darkening" effect when shifting from a major to a minor chord is generally more perceptually salient than the "brightening" *modal* effect postulated by the hypothesis.

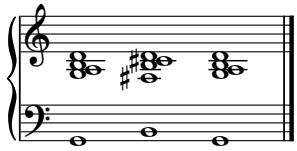


Figure 22: ABA16

15.6.5 Chord tone change

For the *Chord tone change* stimulus set, the mean across all participant groups was 2.12 of 3 possible hypothesis-supporting answers (70.62%). The strongest stimuli in this set consisted of alternations between parallel tonic chords. A stimulus consisting of the chords G–Gm–G (ABA6) was strong in all groups, particularly both groups of musicians (educated musicians 95.9%, hobby/amateur musicians 90.0%, non-musicians 66.7%). A stimulus with the opposite movement (Gm–G–Gm; ABA7) was also strong for both groups of musicians, although, somewhat surprisingly, less strong overall (educated musicians 84.5%, amateur musicians 60.0%, non-musicians 35.4%).

15.6.6 Circle of fifths progressions

Every circle of fifths (COF) stimulus set had a total performance mean of less than 50 percent, making them among the weakest stimulus sets in the experiment. The hypothesis proposed by Mulholland and Hojnacki maintains that number of positions traversed through the COF will influence the luminosity effect: The effect is present already with stepwise movement through the circle of fifths, but successively stronger with movements of two, three, and four positions respectively (Mulholland & Hojnacki, 2013, p. 167).

Did some COF progressions in the experiment seem stronger than others? *COF4* was very weak with a stimulus set total mean of 24.79% hypothesis-supporting answers, so that is already a blow to Mulholland and Hojnacki's hypothesis. On closer inspection, each of the four stimuli had a mean across participant groups of 30.5%, 7.2%, 13.1%, and 48.3% respectively, so the results were not internally consistent either.

COF1 was also weak, and although this corresponds to the hypothesis, the mean across all participant groups was only 0.44 points out of 2 possible (21.82%) and thus did not support the luminosity effect predicted for stepwise movement in the circle of fifths. Looking at the individual scores for the two stimuli in the *COF1* stimulus set revealed something interesting: While the clockwise, stepwise progression of major chords (Figure 23a) was perceived as somewhat brightening in the different participant groups (educated musicians 43.9%, hobby/amateur musicians 32.5%, non-musicians 29.2%), reversing this pattern to a counterclockwise progression (Figure 23b) did not yield comparable results. In the latter stimulus, only 3.4% of educated musicians found the effect darkening. In the hobby/amateur musician group, the number was 0%—the only such case in the entire experiment.[156] In fact, all pairs of stimuli in

156 As a reminder, even though no participants in the AM group chose the hypothesis-supporting answer for this stimulus, it does not mean they chose the hypothesis-conflicting answer either; they may have used the "Vague / No

the COF stimulus sets showed the same trend: The same chord progression pattern performed clockwise and counterclockwise respectively was judged radically differently by all participant groups, and did not yield comparable or consistent results. This was the case for both major chord progression and minor chord progressions. Hypothesis *B* predicted that direction of movement through the COF would influence the perceived sense of harmonic brightening or darkening respectively. While this effect was expected to vary with the number of COF positions traversed—as suggested by Mulholland and Hojnacki—the hypothesis expected it to be internally consistent for *COF1*, *COF2*, *COF3*, and *COF4* respectively. These results suggest that *other* factors than the direction of movement through the COF are more influential on perceived brightness.

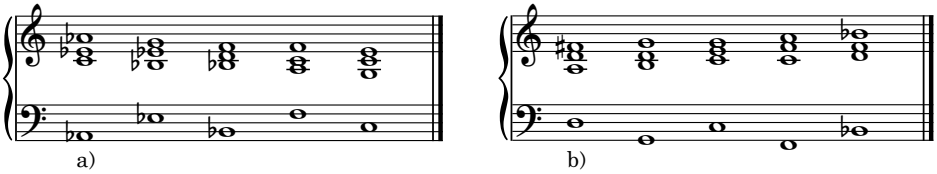


Figure 23: Grade7 (a) and Grade7 reverse (b)

Educated musicians had a markedly higher score than hobby/amateur musicians for *COF2* (52.2% vs. 40.62%), but for *COF3* the results were almost exactly reversed (40.04% vs. 51.88%). As seen from the numbers, the best performing group in both stimulus sets was in the vicinity of 50% hypothesis-supporting answers. These results do not seem to convincingly support hypothesis *B*: *Clockwise modulations along the circle of fifths generally create a brightening effect. Counterclockwise modulations along the circle of fifths generally create a darkening effect.*

15.6.7 Other/mixed

For the *Other/mixed* stimulus set, the mean across all participant groups was 3.61 of 5 possible hypothesis-supporting answers (72.29%), making this set the highest-scoring in the experiment. While all five stimuli in the set had a relatively high score for educated musicians, one stimulus in particular stood out: Grade9 (Figure 24) had the highest mean score for each participant group in the entire experiment, with 97.3% for educated musicians, 92.5% for hobby/amateur musicians, and 89.6% for non-musicians. Taken together with some of the results in the *modal mixture* stimulus set, this could suggest that subdominants are generally the most salient functions for harmonic luminosity effects, whenever they appear with the opposite quality of the tonic. However, ABA14 (see above) is a notable exception to this trend,

clear change either way" reply option to a certain degree.

making the results inconclusive. In any case, it is difficult to conclude whether this has to do with the subdominant function itself or is mainly caused by the juxtaposition of major and minor chords (major tonic and minor subdominant, or minor tonic and major subdominant).

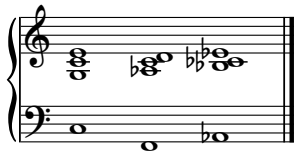


Figure 24: Grade9

16 Discussion

16.1 Conclusions and implications

The results of the experiment were conflicting and inconclusive. The overall performance of participants did not yield strong support for the perceptual salience of harmonic luminosity. Even when isolating the group of current or former students in higher music education, the total performance mean of 51.29% hypothesis-supporting answers was rather unconvincing. On the other hand, participants with musical background (both educated and amateur musicians) did score higher than non-musicians, and the difference was statistically significant. This suggests that the perception of harmonic luminosity might require some degree of perceptual learning. As discussed in Chapter 14, this is to be expected, because a correspondence between musical harmony and visual light does not easily fit into any of Spence's (2019) four classes of explanation for cross-modal correspondences: statistical, structural, semantic/linguistic, or emotionally mediated. Since musical harmony is a composite perceptual phenomenon loaded with ambiguous musical, cultural, and semiotic meanings, any potential cross-modal correspondence between musical harmony and visual light could be expected to be either a) founded on less one-dimensional structures than those suggested by Spence, b) associated with certain forms of perceptual expertise, or c) both.

However, the difference between musicians and non-musicians could also stem from different levels of musical ability (e.g., the capacity to distinguish aurally between different musical qualities in the listening tasks), or the intellectual knowledge of top-down concepts (e.g., that minor chords are "dark" relative to major chords, that major-seventh chords are "bright", etc.). Thus, the possibility must be considered that the relatively lower performance by non-musicians is indicative of a generally lower level of musical sophistication—and may have nothing to do with the salience of harmonic luminosity. Such questions, and the role of perceptual learning, should be addressed in future research on the topic.

The principle of harmonic luminosity offers two conceptualizations of the harmonic mechanics governing brightening/darkening effects:

A) *Adding sharps to (or subtracting flats from) a mode makes it brighter. Adding flats (or subtracting sharps) makes it darker.*

B) *Clockwise modulations along the circle of fifths generally create a brightening effect. Counterclockwise modulations along the circle of fifths generally create a darkening effect.*

While the two conceptualizations fundamentally express the same underlying harmonic procedure, they were treated as two separate hypotheses, and two separate sets of tasks were constructed in the experiment: Session 1 in accordance with the hypothesis *A*, Session 2 in accordance with hypothesis *B*. Research question 1 was: *Do the two hypotheses consistently predict the participants' judgement of the stimuli?* A significant difference in performance between the two sessions of the experiment was not expected. However, this was indeed the case, with Session 1 achieving a higher percentage of hypothesis-supporting answers than Session 2. This is in accordance with the generally weak and self-contradictory results of Session 2, discussed in the Results section. Direction of movement through the COF did not conclusively or consistently predict the judgement of harmonic brightening or darkening. The mean performance of Session 1, however, was not convincing either, with 52.1 percent hypothesis-supporting answers across all participants. On an overall level, then, the answer to RQ1 is that the hypotheses did not consistently predict the participants' judgments, although there was more support for the hypothesis *A* (adding/subtracting accidentals) than for hypothesis *B* (movement through the circle of fifths).

The second research question was: *How are the participants' performances influenced by musical experience, main instrument, genre, and other factors?* Again, the results showed a statistically significant effect of musical background. This effect applied to participants formerly or currently in higher music education as well as to amateur/hobby musicians, and the two groups performed at a very similar level. On the other hand, performance was not shown to be statistically influenced by any other background factors.

Breaking down the overall results by grouping the stimuli according to harmonic themes and separating the participants by musical background, some nuances started to appear. Again, there were some differences in luminosity judgement in COF movement by different number of positions, but not in the systematic and successively stronger fashion proposed by Mulholland and Hojnacki. None of the participants groups performed predictably on any of the stimuli in the COF set, and there was no consistency in the judgement of structurally identical progressions played in opposite directions (clockwise and counterclockwise respectively). In terms of consistency, there was no difference between major chord progressions and minor chord progressions. Neither was the luminosity effect successively stronger with number of positions traversed, like Mulholland and Hojnacki's hypothesis predicted. *COF4*, which was expected to have the strongest luminosity effect of all the COF progressions, had a mean of only 24.79 percent hypothesis-supporting answers. In conclusion, this study found

no empirical support whatsoever for the hypothesis that clockwise movement through the circle of fifths induces a sensation of harmonic brightening while counterclockwise movement elicits harmonic darkening.

For chromatic mediants, the results were too inconsistent to draw conclusions either way. One chromatic mediant, the ♭VI chord in a major key context, was found to be strongly brightening by all participant groups. Furthermore, this effect was unlikely to be influenced by pitch contour. Nonetheless, this result was based on a single stimulus, and should—given the generally inconsistent results in the study—be viewed with some reservation.

For stimuli using modal mixture chords, there was some—albeit not conclusive—support for the luminosity principle. Subdominants of the opposite quality from the tonic (i.e., a minor subdominant juxtaposed with a major tonic, or a major subdominant juxtaposed with a minor tonic) were generally judged strongly in accordance with hypothesis *A* by participants with musical background, although one stimulus complicated the picture. This exception stimulus could possibly be explained by ascending (and thus "brightening") pitch contour in both top voice and bass overriding the supposed harmonic darkening. If this stimulus were considered an outlier due to these circumstances, there would be some support for proposing that borrowed subdominant chords generally constitute a harmonic brightening or darkening in a chord progression.

The strongest luminosity effect in the study was found in the themed stimulus set containing chord tone changes. Shifts from a major triad to a minor triad on the same root, and vice versa, were overwhelmingly judged by both educated musicians and amateur musicians to be brightening/darkening in accordance with hypothesis *A*.[157] The notion that "major chords are bright, minor chords are dark"[158] was possibly active in the judgement of other stimuli as well, even in situations where this contradicted the hypothesis. This was seen most clearly in the stimulus consisting of the chords G^{ADD9} Bm^{ADD9} G^{ADD9}, in which the C♯ ninth introduced by the B minor chord constitutes a clear brightening of the G major mode, according to the harmonic luminosity principle. However, this chord was judged to be brightening by only 18.6 percent of participants. This could hardly be explained by any other factor than it being a minor (and thus "dark") chord juxtaposed with a major (and thus "bright") chord.

These results allow us to address research questions 3 and 4. RQ3 was: *What types of harmonic movement/juxtaposition cause the greatest luminosity effect, according to participants' judgements?* While some of the results were inconclusive and, to some degree, self-contradictory,

157 Incidentally, this is the same effect highlighted by Riemann (See the extract in Chapter 13.1).

158 Cf. the "major–light, minor–dark" association found in 92 percent of adults by Bonetti and Costa (2019).

there was certain support for the luminosity effect of borrowed subdominant chords juxtaposed with a tonic chord. Also, the juxtaposition of major and minor chords on the same root seemed to induce a consistent and hypothesis-supporting luminosity effect in the groups of participants with musical background. RQ4 was: *In the circle of fifths, which number of positions traversed causes the greatest luminosity effect, according to participants' judgements?* This question was ultimately deemed irrelevant, because direction of movement in the COF was not found to predict judgement of luminosity in the first place.

In conclusion, the study found limited support for the general principle of harmonic luminosity. On the other hand, there was no clear support for refuting the principle either. Some harmonic phenomena had stronger results than others. Also, the statistically significant effect of musical background indicated that some perceptual learning is involved. It is conceivable that a study where music students were divided into an experimental group and a control group would yield stronger results for the group that had been put through a training session. Thus, it seems premature to discard harmonic luminosity as a methodological tool in aural training on the basis of the present study, as such exposure might itself constitute the training necessary for accommodating harmonic luminosity as a perceptual schema. More research is needed in order to address this issue.

The main takeaway of the study is that the association between musical harmony and brightness is probably more multifaceted and complex than is often portrayed in the literature. Other factors than those proposed by the harmonic luminosity principle appear to influence this association: Pitch contour (particularly in the top voice) is likely to play a significant role, and major and minor chords appear to have a certain intrinsic brightness/darkness effect regardless of their harmonic role. It is also worth reminding that in this study, the influence of timbral effects on luminosity judgement was largely controlled through experiment design. In actual music, such effects are likely significant.

16.2 Possible reasons for non-confirmation

There are several possible reasons why the study was not able to find convincing support for the harmonic luminosity principle. The most relevant will be discussed here.

16.2.1 Flawed principle/hypotheses

It is possible that the concept of harmonic luminosity itself is weak, with regard to its perceptual basis. That is, it may postulate a systematic association between musical harmony and visual brightness that most listeners do not experience or agree with. However, this does not mean that harmonic luminosity is not both systematic and perceptually salient for those familiar with the concept. Conceptual familiarity (cf. perceptual learning) might indeed be a prerequisite for this perceptual salience: It is likely that listeners unfamiliar with the concept generally have a rather rudimentary awareness, if any, of their associations between musical harmony and visual brightness. Thus, this study should not be held up as evidence that proponents of harmonic luminosity are wrong. It is, first and foremost, a clarification that the perception of harmonic luminosity is less intuitive than is often (tacitly) presupposed. In addition, it is likely influenced by other, contextual parameters usually not problematized in the literature.

This argument also comes in a softer version: It is possible that the harmonic luminosity *principle* proposed in this study is flawed.[159] This principle was based strictly on mentions of harmonic brightness and darkness in the reviewed literature of Part II of this dissertation, and it formed a reasonably coherent theory of the concept. Therefore, I find it unlikely to be a misrepresentation of the authors' intent. However, it is possible that the generalization of specific harmonic instances into a generative principle (i.e., the harmonic luminosity principle) is a methodological overreach.

It is also worth repeating here that Huron (2006) has proposed a partly conflicting hypothesis of chromatic mediants and luminosity. While his hypothesis, too, finds little support in the present study, it is a reminder that there are indeed conflicting views about brightening and darkening harmony, even among musically sophisticated listeners. The experience of brightness and darkness in musical harmony, then, might simply be much more complex and multifaceted than proponents of harmonic luminosity have taken into account. For example, the concept of "brightness" itself has several connotations besides visual light. Eitan et al. point out how this can lead to confounding metaphorical dimensions:

> Indeed, in a recent study of metaphorical descriptions of music (Eitan, Katz, & Shen, 2010) factor analysis revealed that apparently synaesthetic metaphors for sound, such as light/dark or sweet/sour, were grouped together with emotional terms like happy/sad in a valence factor. Thus, cross-domain mappings may, even for sighted individuals, shed their sensory origins, as a domain originally serving

159 The harmonic luminosity principle: *In general, major or augmented intervals above a mode's root create brightness, relative to minor or diminished intervals.*

> as a target domain (emotion) becomes the new source domain. Thus, when sound is described as "dark," it may not be vision that directly serves as a source domain for a metaphorical description of sound, but "dark" emotion or mood. (Eitan et al., 2012, p. 42)

In this perspective, harmonic luminosity might not be a purely audiovisual correspondence, but one that includes other dimensions of "brightness," e.g., emotion, verticality, or timbre. If such dimensions create independent "brightness effects" that that do not correspond to the principle of harmonic luminosity, it could explain the heterogeneous and often self-contradictory results of this study.

16.2.2 Methodological weaknesses

It is possible that the modest support for harmonic luminosity in this study is not due to flawed hypotheses, but to methodological choices. Some of these choices were deliberate—like limiting the auditory role of timbre and pitch contour—while others were not.

One possibility is that harmonic luminosity is not a "true" cross-modal correspondence, but rather a theoretical, top-down concept with CMC claims. As such, it might be a real phenomenon in the sense that there exists a tradition of perceptual consensus, but that this consensus is contingent on a particular listener intention, learning the theoretical concept first, etc. As mentioned, Spence (2011) suggests that CMCs are generally shared by a large number of people, or even universal. Based on the results of the present study, this is clearly not the case with harmonic luminosity. This indicates that the phenomenon is not learned purely through intuition. Thus, if a systematic mapping between musical harmony and visual brightness is unlikely to develop independently in any individual from sheer environmental exposure to these perceptual categories—without learning a theoretical concept—then it might not be fitting to categorize it as a CMC, according to the commonly accepted understanding of such phenomena. This matter is of less importance for harmonic luminosity's relevance to aural training, which in any case would need to include some degree of learning. However, it could have methodological ramifications: While the present study did not use speeded classification tasks, the paradigmatic experimental setup in CMC research for recent decades (Parise & Spence, 2012), it still relied on spontaneous sensory cue integration. If harmonic luminosity is *not* based on spontaneous sensory cue integration, however—but rather on a theoretical, top-down concept—a different experimental design than the one chosen here should be contemplated. It is also possible that harmonic luminosity should be considered a purely endo-musical phenomenon, and that references to "brightness" and "darkness" are not

symptomatic of real correspondence with the visual modality. It is unclear how this would affect the experimental design.

Furthermore, the way the stimuli were constructed may have impacted the results significantly. Both the pilot study and several individual listening tasks in this study point to the strong influence of pitch contour. There is a striking discrepancy between the strong results of the pilot study (in which pitch contour was *not* controlled) and the weaker results of the final study (in which pitch contour *was* controlled). Given this discrepancy, it is plausible that aligning the brightness/darkness effects stemming from harmonic movement with those stemming from pitch contour (i.e., making sure harmony and pitch contour "pull in the same direction") might substantially have changed the outcome of the study. However, as discussed in Chapter 14, I find this to be a biased and impaired approach, and the results of such a study would be of limited value in furthering our understanding of harmonic luminosity. It is possible that the pitch–brightness correspondence is generally more perceptually salient than a harmony–brightness correspondence. While this would be a setback to the plausibility of the harmonic luminosity principle, it would also be an important clarification. Thus, a possibility for future research is to control the effect of pitch contour in an even more systematic manner, e.g., by using two versions of every harmonic stimulus: one with supporting and one with contrasting pitch contour.

An alternative approach for diminishing the effect of pitch contour, which also addresses another possible methodological weakness in the present study, is to focus more on harmonic modulations than on harmonic progressions. (For instance, Mulholland and Hojnacki's theory of COF movement and luminosity is, strictly speaking, concerned with *modulations* and not progressions.) The experimental setup mostly used chord juxtapositions rather than traditional harmonic *progressions*, and these were played back at a moderate tempo. Thus, the difference between progressions and modulations in this case is more theoretical than it is perceptual. Still, the setup consisted primarily of chord-to-chord movement in relatively short snippets, without much harmonic context. The lack of a clear tonality may have emphasized the perceptual salience of brightness/darkness effects stemming from pitch contour, at the expense of *harmonic* luminosity, especially if the latter were to be contingent on a more established sense of tonality. Thus, it is conceivable that harmonic luminosity would obtain a relatively higher perceptual salience if "luminosity chords" were used more sparingly or introduced through clear modulatory processes, e.g., by allowing them to be preceded by a temporally prolongated state of diatonicism. Using more tonal context might also eliminate the possible perceptual duality of chords that can function as either chromatic mediants or secondary dominants, and which might be judged differently in these cases with regard to luminosity.

16.2.3 Technical factors

In the study, 139 participants used a computer, 95 used a mobile phone, and two used a tablet. The experiment was technically suited for all these devices (as attested by Gorilla Experiment Builder and confirmed in trial runs), but there was no control of participants' technical devices (e.g., use of headphones) and listening environment. While participants were encouraged to use headphones, this was not controlled by the experimental setup. In retrospect, implementing a headphone check would have been both prudent and feasible. In a recent study on online data collection methods in auditory perception research, Eerola et al. (2021) offer some insight into how such tests—or the lack thereof—might influence results. In the 12 peer-reviewed studies that have so far implemented a headphone check developed by Woods et al. (2017), the mean and median failure rates[160] are 16.5% and 17.2% respectively (with a range of 0–40%). A different study, using a headphone check developed by Milne et al. (2021), found a failure rate of 49% (Eerola et al., 2021, p. 10). In short, it cannot be ruled out that the present study was impacted by a low use of headphones and/or noisy listening environments. On the other hand, one may once again point to ecological validity and question whether harmonic luminosity is a valuable tool in aural training if it is contingent on highly specific listening conditions.

16.3 Future research

This is, to the best of my knowledge, the first empirical study on harmonic luminosity perception. As such, it was never to be regarded as the final word on the matter—whatever results it was to find. The concept of brightening and darkening harmony—described in theoretical literature for decades—remains common in harmonic discourse and has gained traction following the popularization and democratization of music theory through social media. While certain aspects of the harmonic luminosity principle tested in this study found no empirical support, other aspects were more positive. The picture is not black and white. Harmonic luminosity's relevance as a component in harmonic aural training is dubious, based on the results found here. But even so, the concept describes a dimension of harmonic experience that merits attention in and of itself—whether or not it is a "true" cross-modal correspondence.

One of the most pressing questions is whether harmonic luminosity can reasonably be described through a coherent, systematic theory (as was attempted in this study), or if it is largely contextual. More theoretical work is needed to improve its conceptual framework, examining

160 That is, failure of the headphone check.

how it is applied in practical music theory discourse, and determining to which degree general principles can be extrapolated from singular descriptions of brightening or darkening harmonic processes. Answering questions like these would be central to determining harmonic luminosity's future status as a research object.

If harmonic luminosity is to remain a relevant concept in music theory, more empirical studies are also needed. Improved and more fitting experiment designs should be pursued. These might include stimuli with more harmonic context, for instance clearer modulations. The role of perceptual learning should also be explored, for example through the use of treatment and control groups. A somewhat more controlled listening environment in online studies can be achieved through simple technological means, such as headphone checks.

The biggest challenge appears to be the perceptual role of pitch contour, and finding some way of minimizing, controlling, or measuring its influence on harmonic luminosity is imperative. One possibility is to contemplate the use of a Shepard tone generator. This option was rejected in the present study for reasons of ecological validity, since one aim of the study was to examine the relevance of harmonic luminosity for aural training pedagogy. Without the concern for a natural classroom setting, future studies might see Shepard tones as a practical way of circumventing the problem of pitch contour altogether. Another possibility is to control pitch contour more systematically through carefully designed stimuli, e.g., pairs of stimuli with luminosity supporting and luminosity contrasting pitch contour respectively.

Dissertation conclusions

This dissertation has been a broad investigation of different notions of "holistic" harmonic listening. Its main contribution has been the elucidation of the concept of "Gestalt listening," originally described, but not clearly defined, by Gary S. Karpinski (2000). Although this concept has engendered curiosity in the aural training community (Chenette et al., 2021; Jarvis, 2015), it has been criticized for being too vague to have real pedagogical implications (Chenette et al., 2021). It is my hope that the conceptual framework offered in this thesis will reawaken an interest in holistic approaches to harmonic listening, from both researchers and aural training instructors. In both fields, further explorations may be carried out in combination with existing trends: In empirical research, the role of embodied cognition and multimodality in music perception, and in music theory pedagogy, the growing interest in harmonic schemata.

The dissertation has also examined the role of metaphor in harmonic perception. While important questions remain unanswered regarding the concept of "brightness" and "darkness" in musical harmony (i.e., *harmonic luminosity*), its ubiquity in the general music theory discourse justifies further research using a variety of methodological approaches.

References

Acevedo, S. (2020). *Harmonic schemata of popular music: An empirical investigation of analytical patterns and their mental representations* (Publication Number 27744507) [Doctoral dissertation, Yale University]. ProQuest Dissertations.

Agawu, V. K. (1991). *Playing with signs: A semiotic interpretation of classic music*. Princeton University Press.

Andrade, C. (2018). Internal, external, and ecological validity in research design, conduct, and evaluation. *Indian journal of psychological medicine, 40*(5), 498–499. https://doi.org/10.4103/IJPSYM.IJPSYM_334_18

Andrianopoulou, M. (2019). *Aural education: Reconceptualising ear training in higher music learning*. Routledge.

Antovic, M. (2009). Musical metaphors in Serbian and Romani children: An empirical study. *Metaphor and Symbol, 24*(3), 184–202. https://doi.org/10.1080/10926480903028136

Anwyl-Irvine, A. L., Massonnié, J., Flitton, A., Kirkham, N., & Evershed, J. K. (2020). Gorilla in our midst: An online behavioral experiment builder. *Behavior Research Methods, 52*(1), 388–407.

Aristotle. (2020). *The Nicomachean ethics* (W. D. Ross, Trans.; L. Brown, Ed.; Revised ed.). Oxford University Press. (Original work published ca. 330 BCE)

Athanasopoulos, G., Eerola, T., Lahdelma, I., & Kaliakatsos-Papakostas, M. (2021). Harmonic organisation conveys both universal and culture-specific cues for emotional expression in music. *Plos one, 16*(1), e0244964.

Austbø, H. (2015). Visualizing visions: The significance of Messiaen's colours. *Music & Practice, 2*. www.musicandpractice.org/volume-2/visualizing-visions-the-significance-of-messiaens-colours/

Bach, J. S. (1983). Koopman: Das Wohltemperierte Klavier [CD]. Erato Classics.

Bach, J. S. (1984). Schiff: Das Wohltemperierte Klavier I [CD]. Decca.

Bach, J. S. (1990). Gieseking: Das Wohltemperierte Klavier I & II [CD]. Deutsche Grammophon.

Barnard, P. J. (2012). What do we mean by the meanings of music? *Empirical Musicology Review, 7*(1–2), 69–80. https://doi.org/https://doi.org/10.18061/1811/52982

Beckerman, M. (1994). *Janáček as theorist*. Pendragon Press.

Blix, H. S. (2013). Learning strategies in ear training. In I. E. Reitan, A. K. Bergby, V. C. Jakhelln, G. Shetelig, & I. F. Øye (Eds.), *Aural perspectives: On musical learning and practice in higher music education* (pp. 97–115). Norwegian Academy of Music / Akademika.

Bonacchi, S. (2015). Some preliminary remarks about the use of the expression "Gestalt" in the scientific debate. *Dialogue and Universalism, XXV*(4), 11–20.

Bonetti, L., & Costa, M. (2019). Musical mode and visual-spatial cross-modal associations in infants and adults. *Musicae Scientiae, 23*(1), 50–68. https://doi.org/10.1177/1029864917705001

Bowen, G. A. (2009). Document analysis as a qualitative research method. *Qualitative research journal, 9*(2), 27–40.

Bregman, A. S. (1990). *Auditory scene analysis: The perceptual organization of sound*. MIT Press.

Bremner, A. J., Caparos, S., Davidoff, J., de Fockert, J., Linnell, K. J., & Spence, C. (2013). "Bouba" and "Kiki" in Namibia? A remote culture make similar shape–sound matches, but different shape–taste matches to Westerners. *Cognition, 126*(2), 165–172. https://doi.org/https://doi.org/10.1016/j.cognition.2012.09.007

Brent, J. (2011). *Modalogy: Scales, modes & chords: The primordial building blocks of music*. Hal Leonard.

Brooks, J. L. (2015). Traditional and new principles of perceptual grouping. In J. Wagemans (Ed.), *The Oxford handbook of perceptual organization* (pp. 57–87). Oxford University Press.

Brower, C. (2000). A cognitive theory of musical meaning. *Journal of Music Theory, 44*(2), 323–379. https://doi.org/10.2307/3090681

Bruce, V., Green, P. R., & Georgeson, M. A. (1996). *Visual perception: Physiology, psychology, and ecology* (3rd ed.). Lawrence Erlbaum Ass.

Brunel, L., Carvalho, P. F., & Goldstone, R. L. (2015). It does belong together: Cross-modal correspondences influence cross-modal integration during perceptual learning. *Frontiers in Psychology, 6*(358). https://doi.org/10.3389/fpsyg.2015.00358

Buccella, A. (2021). The problem of perceptual invariance. *Synthese, 199*(5), 13883–13905. https://doi.org/10.1007/s11229-021-03402-2

Chemero, A. (2009). *Radical embodied cognitive science*. MIT Press.

Chenette, T. (2021). What are the truly aural skills? *Music theory online, 27*(2). https://doi.org/10.30535/mto.27.2.2

Chenette, T., Phillips, A., & Wood, E. (2021). *Does Gestalt hearing exist?* Proceedings of the Future Directions of Music Cognition International Conference, Ohio State University (online). https://doi.org/10.18061/FDMC.2021.0028

Chow, S. L. (1989). An intentional analysis of "affordance" revisited. *Journal for the Theory of Social Behaviour, 19*(3), 357–365.

Christensen, E. (1996). *The musical timespace: A theory of music listening*. Aalborg University Press.

Clarke, E. F. (2005). *Ways of listening: An ecological approach to the perception of musical meaning*. Oxford University Press.

Clarke, E. F. (2012). What's going on: Music, psychology, and ecological theory. In M. Clayton, T. Herbert, & R. Middleton (Eds.), *The cultural study of music: A critical introduction* (2nd ed., pp. 333–342). Routledge.

Clarke, E. F. (2017). Music's meanings. In R. Ashley & R. Timmers (Eds.), *The Routledge companion to music cognition* (pp. 523–533). Routledge.

Cleland, K. D., & Dobrea-Grindahl, M. (2013). *Developing musicianship through aural skills: A holistic approach to sight singing and ear training*. Routledge.

Clendinning, J. P., & Marvin, E. W. (2016). *The musician's guide to theory and analysis* (3rd ed.). W. W. Norton.

Cohn, R. (2011). Tonal pitch space and the (neo-)Riemannian Tonnetz. In E. Gollin & A. Rehding (Eds.), *The Oxford handbook of neo-Riemannian music theories*. Oxford University Press.

Cohn, R., Hyer, B., Dahlhaus, C., Anderson, J., & Wilson, C. (2001). Harmony. *Grove Music Online*. Retrieved February 20, 2022, from https://www.oxfordmusiconline.com/grovemusic/view/10.1093/gmo/9781561592630.001.0001/omo-9781561592630-e-0000050818

Coker, J., Knapp, B., & Vincent, L. (1997). *Hearin' the changes: Dealing with unknown tunes by ear*. Advance Music.

Collier, W. G., & Hubbard, T. L. (2004). Musical scales and brightness evaluations: Effects of pitch, direction, and scale mode. *Musicae Scientiae, 8*(2), 151–173. https://doi.org/10.1177/102986490400800203

Cook, N. (1998). *Music: A very short introduction*. Oxford University Press.

Cook, N. D., & Hayashi, T. (2008). The psychoacoustics of harmony perception: Centuries after three-part harmony entered Western music, research is starting to clarify why different chords sound tense or resolved, cheerful or melancholy. *American Scientist, 96*(4), 311–319.

Cooke, D. (1959). *The language of music*. Oxford University Press.

Cuskley, C., Simner, J., & Kirby, S. (2017). Phonological and orthographic influences in the bouba–kiki effect. *Psychological research, 81*(1), 119–130.

Ćwiek, A., Fuchs, S., Draxler, C., Asu, E. L., Dediu, D., Hiovain, K., Kawahara, S., Koutalidis, S., Krifka, M., Lippus, P., Lupyan, G., Oh, G. E., Paul, J., Petrone, C., Ridouane, R., Reiter, S., Schümchen, N., Szalontai, Á., Ünal-Logacev, Ö., Zeller, J., Perlman, M., & Winter, B. (2022). The *bouba/kiki* effect is robust across cultures and writing systems. *Philosophical Transactions of the Royal Society B: Biological Sciences, 377*(1841), 20200390. https://doi.org/doi:10.1098/rstb.2020.0390

Davies, S. (1994). *Musical meaning and expression*. Cornell University Press.

De Vaus, D. (2014). *Surveys in social research* (6th ed.). Routledge.

DeBellis, M. (1995). *Music and conceptualization*. Cambridge University Press.

DeBellis, M. (2005). Conceptual and nonconceptual modes of music perception. *Postgraduate Journal of Aesthetics, 3*(2), 45–61. https://philpapers.org/rec/DEBCAN

Deliege, I. (1987). Grouping conditions in listening to music: An approach to Lerdahl & Jackendoff's grouping preference rules. *Music Perception: An Interdisciplinary Journal, 4*(4), 325–359. https://doi.org/10.2307/40285378

Denham, S. L., & Winkler, I. (2015). Auditory perceptual organization. In J. Wagemans (Ed.), *The Oxford handbook of perceptual organization* (pp. 601–620). Oxford University Press.

DeNora, T. (2000). *Music in everyday life*. Cambridge-Obeikan.

Deutsch, D. (1982). Organizational processes in music. In M. Clynes (Ed.), *Music, mind, and brain: The neuropsychology of music* (pp. 119–136). Springer US. https://doi.org/10.1007/978-1-4684-8917-0_6

Deutsch, D. (2013). *The psychology of music* (3rd ed.). Academic Press.

Dibben, N. (2001). What do we hear, when we hear music?: Music perception and musical material. *Musicae Scientiae, 5*(2), 161–194.

Dobbins, B. (1986). *Jazz arranging and composing: A linear approach*. Advance Music.

Doll, C. (2017). *Hearing harmony: Toward a tonal theory for the rock era*. University of Michigan Press.

Dolscheid, S., Shayan, S., Majid, A., & Casasanto, D. (2013). The thickness of musical pitch: Psychophysical evidence for linguistic relativity. *Psychological Science, 24*(5), 613–621. https://doi.org/10.1177/0956797612457374

Drabkin, W. (2001). Neapolitan sixth chord. *Grove Music Online*. Retrieved March 14, 2022, from https://www.oxfordmusiconline.com/grovemusic/view/10.1093/gmo/9781561592630.001.0001/omo-9781561592630-e-0000019658

Drouzas, C., & Saitis, C. (2020). Verbal description of musical brightness. 2nd International Conference on Timbre, Thessaloniki (online), Greece.

Eberlein, R. (1997). A method of analysing harmony, based on interval patterns or "Gestalten". In M. Leman (Ed.), *Music, Gestalt, and computing* (pp. 225–236).

Eden Ünlü, S., & Ece, A. S. (2019). Reading notation with Gestalt perception principles: Gestalt algı ilkeleri ile notasyon okuma. *Journal of Human Sciences, 16*(4), 1104–1120. https://doi.org/10.14687/jhs.v16i4.5822

Eerola, T., Armitage, J., Lavan, N., & Knight, S. (2021). Online data collection in auditory perception and cognition research: Recruitment, testing, data quality and ethical considerations. *Auditory Perception & Cognition, 4*(3–4), 1–30. https://doi.org/10.1080/25742442.2021.2007718

Ehrenfels, C. v. (1890/1988). On 'Gestalt qualities'. In B. Smith (Ed.), *Foundations of Gestalt theory* (pp. 82–117). Philosophia.

Ehrenfels, C. v. (1932/1988). On Gestalt qualities. In B. Smith (Ed.), *Foundations of Gestalt theory* (pp. 121–123). Philosophia.

Eichert, R., Schmidt, L., & Seifert, U. (1997). Logic, gestalt theory, and neural computation in research on auditory perceptual organization. In M. Leman (Ed.), *Music, Gestalt, and computing* (pp. 70–88).

Eitan, Z. (2013). Musical objects, cross-domain correspondences, and cultural choice: Commentary on "Cross-cultural representations of musical shape" by George Athanasopoulos and Nikki Moran. *Empirical Musicology Review, 8*(3–4), 204–207.

Eitan, Z. (2017). Musical connections: Cross-modal correspondences. In R. Ashley & R. Timmers (Eds.), *The Routledge companion to music cognition* (pp. 213–224). Routledge.

Eitan, Z., & Granot, R. Y. (2006). How music moves: Musical parameters and listeners' images of motion. *Music Perception, 23*(3), 221–248.

Eitan, Z., Katz, A., & Shen, Y. (2010). Effects of pitch register, loudness and tempo on children's use of metaphors for music. In *Poster presentation to the 11th International Conference of Music Perception and Cognition (ICMPC11), Seattle, Washington*.

Eitan, Z., Ornoy, E., & Granot, R. Y. (2012). Listening in the dark: Congenital and early blindness and cross-domain mappings in music. *Psychomusicology, 22*(1), 33–45.

Eitan, Z., & 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*(3), 405–422.

Eitan, Z., Timmers, R., & Adler, M. (2017). Cross-modal correspondences and affect in a Schubert song. In D. Leech-Wilkinson & H. M. Prior (Eds.), *Music and Shape* (pp. 58–86). Oxford University Press.

Feldman, J. (2015). Bayesian models of perceptual organization. In J. Wagemans (Ed.), *The Oxford handbook of perceptual organization* (pp. 1008–1026). Oxford University Press.

Ferrara, J. M. (2001). *Jazz piano and harmony: An advanced guide*. John Ferrara Music Publishing.

Flores d'Arcais, G. B. (1975). Einflüsse der Gestalttheorie auf die moderne kognitive Psychologie. In S. Ertel, L. Kemmler, & M. Stadler (Eds.), *Gestalttheorie in der Modernen Psychologie: Wolfgang Metzger zum 75. Geburtstag* (pp. 45–57). Steinkopff. https://doi.org/10.1007/978-3-642-72312-4_5

Fraisse, P. (1975). Is rhythm a Gestalt? In S. Ertel, L. Kemmler, & M. Stadler (Eds.), *Gestalttheorie in der Modernen Psychologie: Wolfgang Metzger zum 75. Geburtstag* (pp. 227–232). Steinkopff. https://doi.org/10.1007/978-3-642-72312-4_21

Francis, M. (2021). New aural pedagogy for the twenty-first century: Introducing a course for undergraduate students at the Royal College of Music, London. In K. D. Cleland & P. Fleet (Eds.), *The Routledge companion to aural skills pedagogy* (pp. 210–221). Routledge.

Fryer, L., Freeman, J., & Pring, L. (2014). Touching words is not enough: How visual experience influences haptic–auditory associations in the "Bouba–Kiki" effect. *Cognition, 132*(2), 164–173. https://doi.org/https://doi.org/10.1016/j.cognition.2014.03.015

Fulker, R. (2011). *Steve Reich: 'Composers reflect their time and place'* [Interview]. Deutsche Welle. https://p.dw.com/p/12hNX

Gabrielsson, A., & Juslin, P. N. (2003). Emotional expression in music. In R. J. Davidson, K. R. Scherer, & H. H. Goldsmith (Eds.), *Handbook of affective sciences* (pp. 503–534). Oxford University Press.

Gaver, W. W. (1993a). How do we hear in the world? Explorations in ecological acoustics. *Ecological psychology, 5*(4), 285–313.

Gaver, W. W. (1993b). What in the world do we hear? An ecological approach to auditory event perception. *Ecological psychology, 5*(1), 1–29.

Getz, L. M., & Kubovy, M. (2018). Questioning the automaticity of audiovisual correspondences. *Cognition, 175*, 101–108. https://doi.org/https://doi.org/10.1016/j.cognition.2018.02.015

Gibbs, R. W. (2017). *Metaphor wars: Conceptual metaphors in human life*. Cambridge University Press.

Gibson, E. J. (1969). *Principles of perceptual learning and development*. Appleton-Century-Crofts.

Gibson, J. J. (1966). *The senses considered as perceptual systems*. Houghton Mifflin.

Gibson, J. J. (1979). *The ecological approach to visual perception*. Lawrence Erlbaum.

Gibson, J. J., & Gibson, E. J. (1955). Perceptual learning: Differentiation or enrichment? *Psychological review, 62*(1), 32–41.

Gjerdingen, R. O. (1994). Apparent motion in music? *Music Perception, 11*(4), 335–335. https://doi.org/10.2307/40285631

Gjerdingen, R. O. (2007). *Music in the galant style*. Oxford University Press.

Gjerdingen, R. O., & Perrott, D. (2008). Scanning the dial: The rapid recognition of music genres. *Journal of New Music Research, 37*(2), 93–100. https://doi.org/10.1080/09298210802479268

Godøy, R. I. (2010). Gestural affordances of musical sound. In R. I. Godøy & M. Leman (Eds.), *Musical gestures: Sound, movement, and meaning* (pp. 103–125). Routledge.

Godøy, R. I. (2017). Key-postures, trajectories and sonic shapes. In D. Leech-Wilkinson & H. M. Prior (Eds.), *Music and Shape* (pp. 4–29). Oxford University Press.

Godøy, R. I. (2018a). Motor constraints shaping musical experience. *Music theory online, 24*(3). https://doi.org/10.30535/mto.24.3.8

Godøy, R. I. (2018b). Motormimetic features in musical experience. In P. Veroli & G. Vinay (Eds.), *Music-dance: Sound and motion in contemporary discourse* (pp. 207–221). Routledge. http://hdl.handle.net/10852/63262

Gold, R., & Segal, O. (2017). The bouba–kiki effect and its relation to the Autism Quotient (AQ) in autistic adolescents. *Research in Developmental Disabilities, 71*, 11–17. https://doi.org/https://doi.org/10.1016/j.ridd.2017.09.017

Gold, R., & Segal, O. (2020). The bouba–kiki effect in persons with prelingual auditory deprivation. *Language Learning and Development, 16*(1), 49–60. https://doi.org/10.1080/15475441.2019.1685386

Goldstein, G. (1993). *Jazz composer's companion*. Advance Music.

Goldstone, R. L., & Hendrickson, A. T. (2010). Categorical perception. *WIREs Cognitive Science, 1*(1), 69–78. https://doi.org/https://doi.org/10.1002/wcs.26

Gordon, C. L., Iacoboni, M., & Balasubramaniam, R. (2018). Multimodal music perception engages motor prediction: A TMS study. *Frontiers in neuroscience*, 736. https://doi.org/10.3389/fnins.2018.00736

Gorow, R. (2006). *Hearing and writing music: Professional training for today's musician* (Expanded 2nd ed.). September Publishing.

Grey, J. M., & Gordon, J. W. (1978). Perceptual effects of spectral modifications on musical timbres. *The Journal of the Acoustical Society of America, 63*(5), 1493–1500.

Grove, D. (1972). *Arranging concepts complete: The ultimate arranging course for today's music*. Alfred Pub. Co.

Guberman, S. (2015). On Gestalt theory principles. *Gestalt Theory, 37*(1), 25–44.

Guberman, S. (2016). Gestalt psychology, mirror neurons, and body–mind problem. *Gestalt Theory, 38*(2/3). https://www.academia.edu/download/53440136/Gestalt_Psychology__Mirror_Neurons__and_Body-Mind_Problem..pdf

Guberman, S. (2017). Gestalt theory rearranged: Back to Wertheimer. *Frontiers in Psychology, 8*. https://doi.org/10.3389/fpsyg.2017.01782

Gur, G. (2008). Body, forces, and paths: Metaphor and embodiment in Jean-Philippe Rameau's conceptualization of tonal space. *Music theory online, 14*(1).

Gurwitsch, A. (1979). *Phenomenology and theory of science*. Northwestern University Press.

Haeckel, E. (1866). *Generelle Morphologie der Organismen. Allgemeine Grundzüge der organischen Formen-Wissenschaft, mechanisch begründet durch die von Charles Darwin reformirte Descendenz-Theorie* (Vol. 1). Reimer.

Hamilton-Fletcher, G., Pisanski, K., Reby, D., Stefańczyk, M., Ward, J., & Sorokowska, A. (2018). The role of visual experience in the emergence of cross-modal correspondences. *Cognition, 175*, 114–121. https://doi.org/https://doi.org/10.1016/j.cognition.2018.02.023

Hansen, N. C. (2010/11). The legacy of Lerdahl and Jackendoff's 'A generative theory of tonal music': Bridging a significant event in the history of music theory and recent developments in cognitive music research. *Danish yearbook of musicology, 38*, 34–55.

Harrison, D. (1994). *Harmonic function in chromatic music: A renewed dualist theory and an account of its precedents*. University of Chicago Press.

Harrison, M. (1995). *Contemporary music theory (level two). A complete harmony and theory method for the pop & jazz musician*. Hal Leonard.

Hatfield, G. (2014). Cognition. In L. Shapiro (Ed.), *The Routledge handbook of embodied cognition* (pp. 361–373). Routledge.

Hatten, R. S. (1995). Metaphor in music. In E. Tarasti (Ed.), *Musical signification: Essays in the semiotic theory and analysis of music* (pp. 373–391). De Gruyter Mouton.

Hebb, D. O. (1949). *The organization of behavior: A neuropsychological theory*. Wiley.

Hedger, S. C. V., Heald, S. L. M., & Nusbaum, H. C. (2015). The effects of acoustic variability on absolute pitch categorization: Evidence of contextual tuning. *The Journal of the Acoustical Society of America, 138*(1), 436–446. https://doi.org/10.1121/1.4922952

Heft, H. (2001). *Ecological psychology in context: James Gibson, Roger Barker, and the legacy of William James's radical empiricism*. Lawrence Erlbaum.

Hendrickson, A., Kachergis, G., Gureckis, T., & Goldstone, R. (2010). Is categorical perception really verbally mediated perception? In *Proceedings of the Annual Meeting of the Cognitive Science Society* (Vol. 32, No. 32).

Herbst, A. (2021). The tortoise and the magic tree: Strategies to develop comprehensive and holistic music analytical listening skills through the use of 'ghost scores'. In K. D. Cleland & P. Fleet (Eds.), *The Routledge companion to aural skills pedagogy* (pp. 190–209). Routledge.

Hindle, D., & Godsil, S. (2006). The song of the siren: Some thoughts on idealization and creativity in Martinů's *Julietta*. *The International Journal of Psychoanalysis, 87*(4), 1087–1102. https://doi.org/10.1516/BB2H-BQXY-GEWW-88YW

Hovland, E. (2009). In search of lost hearing: On the concept of "musikalisches Hören" and the romantic paradigm of music. *Studia musicologica Norvegica, 35(2009)*, 177–195.

Hughes, B., & Lavengood, M. (2021). Introduction to harmonic schemas in pop music. In *Open music theory (Version 2)*. https://viva.pressbooks.pub/openmusictheory/chapter/intro-to-pop-schemas/

Huron, D. (2006). *Sweet anticipation: Music and the psychology of expectation*. MIT Press.

Huron, D. (2016). *Voice leading: The science behind the musical art*. MIT Press.

Imhof, M. (2020). Psychology. In D. L. Worthington & G. D. Bodie (Eds.), *The handbook of listening*. Wiley Online Library.

Iyer, V. (2002). Embodied mind, situated cognition, and expressive microtiming in African-American music. *Music Perception, 19*(3), 387–414.

Jackendoff, R., & Lerdahl, F. (2006). The capacity for music: What is it, and what's special about it? *Cognition, 100*(1), 33–72. https://doi.org/10.1016/j.cognition.2005.11.005

Jansen, E., & Povel, D.-J. (2004). The processing of chords in tonal melodic sequences. *Journal of New Music Research, 33*(1), 31–48. https://doi.org/10.1076/jnmr.33.1.31.35396

Jarvis, B. E. (2015). Hearing harmony holistically: Statistical learning and harmonic dictation. *Engaging Students: Essays in Music Pedagogy, 3*. Retrieved 24. January, 2022, from http://flipcamp.org/engagingstudents3/essays/jarvis.html

Jimenez, I., & Kuusi, T. (2018a). Connecting chord progressions with specific pieces of music. *Psychology of Music, 46*(5), 716–733.

Jimenez, I., & Kuusi, T. (2018b). What helps jazz musicians name tunes from harmony? The relationship between work with harmony and the ability to identify well-known jazz standards from chord progressions. *Psychology of Music, 48*(2), 215–231. https://doi.org/10.1177/0305735618793005

Jimenez, I., Kuusi, T., Czedik-Eysenberg, I., & Reuter, C. (2021). Identifying songs from their piano-driven opening chords. *Musicae Scientiae*. https://doi.org/10.1177/10298649211003631

Jimenez, I., Kuusi, T., & Doll, C. (2020). Common chord progressions and feelings of remembering. *Music & Science, 3*. https://doi.org/10.1177/2059204320916849

Jimenez, I., Kuusi, T., & Ojala, J. (2021). Relative salience of chord-type and chord-voicing changes: A two-oddball paradigm. *Psychology of Music*. https://doi.org/10.1177/03057356211055214

Johannessen, K. S. (2013). Tause kunnskaper i nytt lys? Viktige ansatser og nye bøker innenfor feltet. *Norsk filosofisk tidsskrift, 48*(2), 144–154.

Johnson, E. (2013). Practical tools to foster harmonic understanding. *Music Educators Journal, 99*(3), 63–68. https://doi.org/10.1177/0027432112469539

Johnson, M. (1987). *The body in the mind: The bodily basis of meaning, imagination, and reason*. University of Chicago Press.

Johnson, M. L., & Larson, S. (2003). "Something in the way she moves" – Metaphors of musical motion. *Metaphor and Symbol, 18*(2), 63–84. https://doi.org/10.1207/S15327868MS1802_1

Juslin, P. N., & Sloboda, J. A., (Eds.). (2010). *Handbook of music and emotion: Theory, research, applications*. Oxford University Press.

Karlin, F., & Wright, R. (2004). *On the track: A guide to contemporary film scoring* (2nd ed.). Routledge.

Karpinski, G. S. (1990). A model for music perception and its implications in melodic dictation. *Journal of Music Theory Pedagogy, 4*(2), 191–229.

Karpinski, G. S. (2000). *Aural skills acquisition: The development of listening, reading, and performing skills in college-level musicians*. Oxford University Press.

Karpinski, G. S. (2007). *Manual for ear training and sight singing*. W. W. Norton.

Karpinski, G. S. (2021). The seeing ear: Toward a rationale for dictation. In K. D. Cleland & P. Fleet (Eds.), *The Routledge companion to aural skills pedagogy* (pp. 37–46). Routledge.

Kastner, M. P., & Crowder, R. G. (1990). Perception of the major/minor distinction: IV. Emotional connotations in young children. *Music Perception: An Interdisciplinary Journal, 8*(2), 189–201. https://doi.org/10.2307/40285496

Kim, J. N., Large, E. W., Gwon, Y., & Ashley, R. (2018). The online processing of implied harmony in the perception of tonal melodies: Effects of harmonic expectations. *Music Perception, 35*(5), 594–606. https://doi.org/10.1525/mp.2018.35.5.594

Klapetek, A., Ngo, M. K., & Spence, C. (2012). Does crossmodal correspondence modulate the facilitatory effect of auditory cues on visual search? *Attention, Perception, & Psychophysics, 74*(6), 1154–1167.

Klapp, S. T., & Jagacinski, R. J. (2011). Gestalt principles in the control of motor action. *Psychological bulletin, 137*(3), 443–462.

Klapp, S. T., Nelson, J. M., & Jagacinski, R. J. (1998). Can people tap concurrent bimanual rhythms independently? *Journal of Motor Behavior, 30*(4), 301–322.

Klonoski, E. (2000). A perceptual learning hierarchy: An imperative for aural skills pedagogy. *College Music Symposium, 40*. https://symposium.music.org/index.php/40/item/2172-a-perceptual-learning-hierarchy-an-imperative-for-aural-skills-pedagogy?highlight=WyJrbG9ub3NraSJd

Köhler, W. (1929). *Gestalt psychology*. Liverlight.

Korsakova-Kreyn, M., & Dowling, W. J. (2012). Emotion in music: Affective responses to motion in tonal space. In E. Cambouropoulos, C. Tsougras, P. Mavromatis, & K. Pastiades (Eds.), *Proceedings: The ICMPC-ESCOM 2012 joint conference—12th biennial international conference for music perception and cognition (Thessaloniki: Aristotle University of Thessaloniki)* (pp. 550–558). http://icmpc-escom2012.web.auth.gr/files/papers/550_Proc.pdf

Korsakova-Kreyn, M., & Dowling, W. J. (2014). Emotional processing in music: Study in affective responses to tonal modulation in controlled harmonic progressions and real music. *Psychomusicology: Music, Mind, and Brain, 24*(1), 4–20.

Kostka, S., Payne, D., & Almén, B. (2018). *Tonal harmony: With an introduction to post-tonal music* (8th ed.). McGraw-Hill Education.

Kozak, M. (2020). *Enacting musical time: The bodily experience of new music*. Oxford University Press.

Krebiehl, A. (2020). The science behind the main wine aromas, explained. *Wine Enthusiast Magazine*. www.winemag.com/2020/05/27/aromas-wine-describe-guide/

Krueger, J. (2011). Doing things with music. *Phenomenology and the Cognitive Sciences, 10*(1), 1–22.

Krueger, J. (2014). Affordances and the musically extended mind. *Frontiers in Psychology, 4*, Article 1003. https://doi.org/10.3389/fpsyg.2013.01003

Kurby, C. A., & Zacks, J. M. (2008). Segmentation in the perception and memory of events. *Trends in cognitive sciences, 12*(2), 72–79.

Kurth, E. (1931). *Musikpsychologie*. Hesse.

Kuusi, T. (2013). Musical training and musical ability: Effects on chord discrimination. *Psychology of Music, 43*(2), 291–301. https://doi.org/10.1177/0305735613511504

Lahdelma, I., & Eerola, T. (2015). Theoretical proposals on how vertical harmony may convey nostalgia and longing in music. *Empirical Musicology Review, 10*(3), 245–263. http://emusicology.org/article/view/4534/4151

Lahdelma, I., & Eerola, T. (2016). Single chords convey distinct emotional qualities to both naïve and expert listeners. *Psychology of Music, 44*(1), 37–54. https://doi.org/10.1177/0305735614552006

Laitz, S. G. (2008). *The complete musician: An integrated approach to tonal theory, analysis, and listening* (2nd ed.). Oxford University Press.

Laitz, S. G. (2016). *The complete musician: An integrated approach to theory, analysis, and listening* (4th ed.). Oxford University Press.

Lakoff, G., & Johnson, M. (1980). *Metaphors we live by*. University of Chicago Press.

Langfeldt, V. (2017). «*Real time planking*». *En fenomenologisk undersøkelse av representasjonsformer og assosiasjon i jazzmusikeres harmoniske gehør* [Master's degree, Norwegian Academy of Music]. http://hdl.handle.net/11250/2430956

Larson, S. (2012). *Musical forces: Motion, metaphor, and meaning in music*. Indiana University Press.

Lee, J. (2017). *Interview: Jacob Collier (Part 1)*, YouTube, www.youtube.com/watch?v=DnBr070vcNE.

Lehman, F. (2013). Hollywood cadences: Music and the structure of cinematic expectation. *Music theory online, 19*(4).

Lehman, F. (2018). *Hollywood harmony: Musical wonder and the sound of cinema*. Oxford University Press.

Lehmann, A. C., Sloboda, J. A., & Woody, R. H. (2007). *Psychology for musicians: Understanding and acquiring the skills*. Oxford University Press.

Leman, M., & Maes, P.-J. (2014). The role of embodiment in the perception of music. *Empirical Musicology Review, 9*(3–4), 236–246.

Leman, M., & Schneider, A. (1997). Origin and nature of cognitive and systematic musicology: An introduction. In M. Leman (Ed.), *Music, Gestalt, and computing* (pp. 11–29). Springer.

Lerdahl, F. (1988). Tonal pitch space. *Music Perception*, 315–349.

Lerdahl, F., & Jackendoff, R. (1983). *A generative theory of tonal music*. MIT Press.

Levine, M. (1995). *The jazz theory book*. Cher.

Lovell, J. (2021). Rethinking integration in the music theory curriculum. In K. D. Cleland & P. Fleet (Eds.), *The Routledge companion to aural skills pedagogy* (pp. 79–98). Routledge.

Luchins, A. S. (1975). The place of Gestalt theory in American psychology: A case study. In S. Ertel, L. Kemmler, & M. Stadler (Eds.), *Gestalttheorie in der Modernen Psychologie: Wolfgang Metzger zum 75. Geburtstag* (pp. 21–44). Steinkopff. https://doi.org/10.1007/978-3-642-72312-4_4

Ludwig, V. U., Adachi, I., & Matsuzawa, T. (2011). Visuoauditory mappings between high luminance and high pitch are shared by chimpanzees (*Pan troglodytes*) and humans. *Proceedings of the National Academy of Sciences, 108*(51), 20661. https://doi.org/10.1073/pnas.1112605108

Macpherson, F. (2011). XV—Cross-modal experiences. *Proceedings of the Aristotelian Society, 111*(3_pt_3), 429–468.

Mandler, J. M. (2014). *Stories, scripts, and scenes: Aspects of schema theory*. Psychology Press.

Marks, L. E. (1974). On associations of light and sound: The mediation of brightness, pitch, and loudness. *The American Journal of Psychology, 87*(1/2), 173–188. https://doi.org/10.2307/1422011

Martino, G., & Marks, L. E. (2001). Synesthesia: Strong and weak. *Current Directions in Psychological Science, 10*(2), 61–65. https://doi.org/10.1111/1467-8721.00116

Marvin, E. W., & Brinkman, A. R. (2000). The effect of key color and timbre on absolute pitch recognition in musical contexts. *Music Perception, 18*(2), 111–137.

Mathieu, W. A. (1997). *Harmonic experience: Tonal harmony from its natural origins to its modern expression*. Inner Traditions.

Maurer, D., Pathman, T., & Mondloch, C. J. (2006). The shape of boubas: Sound–shape correspondences in toddlers and adults. *Developmental Science, 9*(3), 316–322. https://doi.org/https://doi.org/10.1111/j.1467-7687.2006.00495.x

Menin, D., & Schiavio, A. (2012). Rethinking musical affordances. *Avant, 3*(2), 201–215.

Merriam-Webster. (n.d.-a). *Chromatic*. Retrieved April 12, 2022 from www.merriam-webster.com/dictionary/chromaticism

Merriam-Webster. (n.d.-b). *Foreshadow*. Retrieved April 13, 2022 from www.merriam-webster.com/dictionary/foreshadow

Merriam-Webster. (n.d.-c). *Somber*. Retrieved April 15, 2021 from www.merriam-webster.com/dictionary/somber

Messiaen, O. (1990). *The technique of my musical language. 1st volume: Text* (J. Satterfield, Trans.). Leduc.

Metz-Göckel, H. (2015). Gestalt theory: History, principle assumptions and research fields. *Dialogue and Universalism, 25*(4), 21–36. https://doi.org/10.5840/du201525482

Metzger, W. (1965). The foundations of artistic experience. *Acta Psychologica, 24*, 409–422. https://doi.org/10.1016/0001-6918(65)90025-9

Meyer, L. B. (1956). *Emotion and meaning in music*. University of Chicago Press.

Milan, E., Iborra, O., de Cordoba, M., Juárez-Ramos, V., Artacho, M. R., & Rubio, J. (2013). The Kiki–Bouba effect: A case of personification and ideaesthesia. *Journal of Consciousness Studies, 20*(1–2), 84–102.

Miller, G. A. (1956). The magical number seven, plus or minus two: Some limits on our capacity for processing information. *Psychological Review, 63*(2), 81–97.

Milne, A. E., Bianco, R., Poole, K. C., Zhao, S., Oxenham, A. J., Billig, A. J., & Chait, M. (2021). An online headphone screening test based on dichotic pitch. *Behavior Research Methods, 53*(4), 1551–1562. https://doi.org/10.3758/s13428-020-01514-0

Mirka, D., (Ed.). (2014). *The Oxford handbook of topic theory*. Oxford University Press.

Mishra, J. (2000). The effect of anxiety on ear training test scores. *Research Perspectives in Music Education, 7*(1), 12–14. www.ingentaconnect.com/content/fmea/rpme/2000/00000007/00000001/art00004

Miyazaki, K. i. (1989). Absolute pitch identification: Effects of timbre and pitch region. *Music Perception, 7*(1), 1–14.

Moldenhauer, H., & Moldenhauer, R. (1978). *Anton Webern: A chronicle of his life and work*. Victor Gollancz.

Molino, J. (1975). Fait musical et sémiologie de la musique. *Musique En Jeu, 17*, 37–62.

Mulholland, J., & Hojnacki, T. (2013). *The Berklee book of jazz harmony*. Berklee Press.

Müllensiefen, D., Gingras, B., Musil, J., & Stewart, L. (2014). The musicality of non-musicians: An index for assessing musical sophistication in the general population. *Plos one, 9*(2), e89642. https://doi.org/10.1371/journal.pone.0089642

Nattiez, J.-J. (1990). *Music and discourse: Toward a semiology of music* (C. Abbate, Trans.). Princeton University Press.

Neely, A. (2016). *Why is major "happy?"*, YouTube, www.youtube.com/watch?v=9rEqrPwVITY.

Noble, W. (1991). Ecological realism and the fallacy of "objectification". In A. Still & A. Costall (Eds.), *Against cognitivism: Alternative foundations for cognitive psychology* (pp. 199–223). Harvester-Wheatsheaf.

Norman, D. (2013). *The design of everyday things* (Revised and expanded ed.). Basic Books.

Norwegian Academy of Music. (2021). *Aural training II*. Norges musikkhøgskole. Retrieved February 14th, 2022 from https://student.nmh.no/en/studyplans/prospectus-2021/courses/aural-training-ii-2

Nussbaum, C. O. (2007). *The musical representation: Meaning, ontology, and emotion*. MIT Press.

Oberman, L. M., & Ramachandran, V. S. (2008). Preliminary evidence for deficits in multisensory integration in autism spectrum disorders: The mirror neuron hypothesis. *Social Neuroscience, 3*(3–4), 348–355. https://doi.org/10.1080/17470910701563681

Ozturk, G., & Kalyoncu, N. (2018). The effect of cooperative learning on students' anxiety and achievement in musical ear training lessons. *Journal of Measurement and Evaluation in Education and Psychology, 9*(4), 356–375. https://doi.org/10.21031/epod.411010

Pallesen, K. J., Brattico, E., Bailey, C., Korvenoja, A., Koivisto, J., Gjedde, A., & Carlson, S. (2005). Emotion processing of major, minor, and dissonant chords: a functional magnetic resonance imaging study. *Annals of the New York Academy of Sciences, 1060*(1), 450–453.

Pando-Naude, V., Patyczek, A., Bonetti, L., & Vuust, P. (2021). An ALE meta-analytic review of top-down and bottom-up processing of music in the brain. *Scientific Reports, 11*(1), 20813. https://doi.org/10.1038/s41598-021-00139-3

Parise, C., & Spence, C. (2013). Audiovisual cross-modal correspondences in the general population. In J. Simner & E. M. Hubbard (Eds.), *The Oxford handbook of synesthesia* (pp. 790–815). Oxford University Press.

Parise, C. V. (2016). Crossmodal correspondences: Standing issues and experimental guidelines. *Multisensory Research, 29*(1–3), 7–28.

Parise, C. V., & Spence, C. (2012). Audiovisual crossmodal correspondences and sound symbolism: A study using the implicit association test. *Experimental Brain Research, 220*(3–4), 319–333.

Parncutt, R. (2009). Prenatal and infant conditioning, the mother schema, and the origins of music and religion. *Musicae Scientiae, 13*(2_suppl), 119–150.

Pease, T., & Pullig, K. (2001). *Modern jazz voicings: Arranging for small and medium ensembles*. Berklee press.

Peebles, C. (2021). Harmonic schemas in aural skills classes. In K. D. Cleland & P. Fleet (Eds.), *The Routledge companion to aural skills pedagogy* (pp. 320–333). Routledge.

Perricone, J. (2018). *Great songwriting techniques*. Oxford University Press.

Persichetti, V. (1961). *Twentieth-century harmony: Creative aspects and practice*. W.W. Norton.

Pinker, S. (2007). *The stuff of thought: Language as a window into human nature*. Penguin Group.

Piston, W. (1987). *Harmony* (5th ed.). W. W. Norton.

Polanyi, M. (1967). *The tacit dimension*. Doubleday.

Pomerantz, J. R., & Cragin, A. I. (2015). Emergent features and feature combination. In J. Wagemans (Ed.), *The Oxford handbook of perceptual organization* (pp. 88–107). Oxford University Press.

Pomerantz, J. R., & Kubovy, M. (1986). Theoretical approaches to perceptual organization: Simplicity and likelihood principles. In *Handbook of perception and human performance, Vol. 2: Cognitive processes and performance.* (pp. 1–46). John Wiley & Sons.

Povel, D.-J., & Jansen, E. (2002). Harmonic factors in the perception of tonal melodies. *Music Perception, 20*(1), 51–85.

Pragglejaz Group. (2007). MIP: A method for identifying metaphorically-used words in discourse. *Metaphor and Symbol, 22*(1), 1–39. https://doi.org/10.1080/10926480709336752

Pratt, G., Henson, M., & Cargill, S. (1998). *Aural awareness: Principles and practice* (Revised ed.). Oxford University Press.

QSR International Pty Ltd. (2020). *NVivo.* In (Version released in March 2020) www.qsrinternational.com/nvivo-qualitative-data-analysis-software/home

Radley, R. (2008). *The real easy ear training book: A beginning/intermediate guide to hearing the chord changes*. Sher Music.

Radocy, R. E., & Boyle, J. D. (2012). *Psychological foundations of musical behavior* (5th ed.). Charles C. Thomas.

Rahn, J., & McKay, J. (1988). The guide-tone method: An approach to harmonic dictation. *Journal of Music Theory Pedagogy, 2*(1), 101–112.

Ramachandran, V. S., & Hubbard, E. M. (2001). Synaesthesia—a window into perception, thought and language. *Journal of Consciousness Studies, 8*(12), 3–34.

Rameau, J.-P. (1971). *Treatise on harmony* (P. Gossett, Ed. & Trans.). Dover Publications. (Original title: *Traité de l'harmonie reduite à ses principes naturels* [Paris, 1722].)

Reed, E. (1991). James Gibson's ecological approach to cognition. In A. Still & A. Costall (Eds.), *Against cognitivism: Alternate foundations for a cognitive psychology* (pp. 171–197). Harvester Wheatsheaf.

Reitan, I. E. (2008). Stress and well-being in the aural training class: The psychological aspect of training for enhanced musician's skills. In M. Hannan (Ed.), *Proceedings of the 17th International Seminar of the Commission for the Education of the Professional Musician (CEPROM), International Society for Music Education (ISME)* (Vol. 85, pp. 85–90).

Reybrouck, M. (1997). Gestalt concepts and music: Limitations and possibilities. In M. Leman (Ed.), *Music, Gestalt, and computing* (pp. 57–69). Springer.

Reybrouck, M. (2005). A biosemiotic and ecological approach to music cognition: Event perception between auditory listening and cognitive economy. *Axiomathes, 15*(2), 229–266.

Reybrouck, M. (2012). Musical sense-making and the concept of affordance: An ecosemiotic and experiential approach. *Biosemiotics, 5*(3), 391–409.

Reybrouck, M. (2015). Music as environment: An ecological and biosemiotic approach. *Behavioral Sciences, 5*(1), 1–26.

Reybrouck, M. (2021). *Musical sense-making: Enaction, experience, and computation*. Routledge.

Riemann, H. (1886). *The nature of harmony* (J. C. Fillmore, Trans.). Theodore Presser.

Riemann, H. (1900). "Variante." In *Musik-Lexikon* (Fünfte vollständig umgearbeitete Auslage ed.). Max Hesse's Verlag.

Rock, I., & Palmer, S. (1990). The legacy of Gestalt psychology. *Scientific American, 263*(6), 84–91.

Rogers, M. (1984). *Teaching approaches in music theory: An overview of pedagogical philosophies*. Southern Illinois University Press.

Rogers, N. (2001). Gary S. Karpinski: "Aural skills acquisition: The development of listening, reading, and performing skills in college-level musicians". *Indiana theory review, 22*(2), 83–93.

Rogers, S. K., & Ross, A. S. (1975). A cross-cultural test of the Maluma–Takete phenomenon. *Perception, 4*(1), 105–106. https://doi.org/10.1068/p040105

Ruud, E. (2020). En økologisk musikkforståelse. In Ø. Varkøy & H. Holm (Eds.), *Musikkfilosofiske tekster. Tanker om musikk – og språk, tolkning, erfaring, tid, klang, stillhet m.m.* Cappelen Damm Akademisk.

Ryd, Y. (2022). *Snø: Reingjeteren Johan Rassa forteller* (H. Sommerfelt, Trans.). Dreyers forlag.

Ryle, G. (1949). *The concept of mind*. Hutchinson.

Sadai, Y. (1980). *Harmony in its systemic and phenomenological aspects*. Yanetz.

Sanders, J. T. (1997). An ontology of affordances. *Ecological psychology, 9*(1), 97–112.

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.

Schaefer, R. S. (2017). Music in the brain: Imagery and memory. In R. Ashley & R. Timmers (Eds.), *The Routledge companion to music cognition* (pp. 25–35). Routledge.

Schaeffer, P. (1977). *Traité des objets musicaux: Essai interdisciplines* (New ed.). Du Seuil.

Schenker, H. (1969). *Five graphic music analyses (Fünf urlinie-tafeln)*. Dover Publications.

Schiavio, A., & Van Der Schyff, D. (2016). Beyond musical qualia. Reflecting on the concept of experience. *Psychomusicology: Music, Mind, and Brain, 26*(4), 366–378. https://doi.org/10.1037/pmu0000165

Schmeling, P. (2011). *Berklee music theory book 2*. Hal Leonard.

Schneider, A. (1997). "Verschmelzung", tonal fusion, and consonance: Carl Stumpf revisited. In M. Leman (Ed.), *Music, Gestalt, and computing* (pp. 115–143). Springer.

Schneller, T. (2013). Modal interchange and semantic resonance in themes by John Williams. *Journal of Film Music, 6*(1), 49–74.

Schoenberg, A. (1954). *Structural functions of harmony* (L. Stein, Ed. Revised 1969 ed.). Ernest Benn.

Schubert, E., & Wolfe, J. (2006). Does timbral brightness scale with frequency and spectral centroid? *Acta Acustica united with Acustica, 92*(5), 820–825.

Schubert, E., Wolfe, J., & Tarnopolsky, A. (2004). Spectral centroid and timbre in complex, multiple instrumental textures. In S. D. Lipscomb, R. Ashley, R. Gjerdingen, & P. Webster (Eds.), *Proceedings of the 8th International Conference on Music Perception & Cognition (Evanston, Illinois)* (pp. 654–657). https://phys.unsw.edu.au/jw/reprints/SchWolTarICMPC8.pdf

Schulte, M., Knief, A., Seither-Preisler, A., & Pantev, C. (2001). Gestalt recognition in a virtual melody experiment. In *Biomag 2000, proceedings of the 12th international conference on biomagnetism* (pp. 107–110). Helsinki University of Technology Espoo, Finland.

Schwartz, A. (2012). *Harmonic brightness and darkness*. Retrieved April 14th, 2021 from https://antonjazz.com/2012/11/brightness-darkness/

Scruton, R. (1983). Understanding music. In *The aesthetic understanding: Essays in the philosophy of art and culture*. Carcanet Press.

Shanahan, D., Broze, Y., & Rodgers, R. (2012). A diachronic analysis of harmonic schemata in jazz. In *Proceedings of the 12th International Conference on Music Perception and Cognition and the 8th Triennial Conference of the European Society for the Cognitive Sciences of Music* (pp. 909–917). https://www.researchgate.net/profile/Daniel-Shanahan-3/publication/307981744_A_Diachronic_Analysis_of_Harmonic_Schemata_in_Jazz/links/57d75f4a08ae5f03b494e575/A-Diachronic-Analysis-of-Harmonic-Schemata-in-Jazz.pdf

Shapiro, L., (Ed.). (2014). *The Routledge handbook of embodied cognition*. Routledge.

Sloboda, J. (2000). *The musical mind: The cognitive psychology of music*. Oxford University Press.

Sloboda, J. A. (2005). *Exploring the musical mind: Cognition, emotion, ability, function*. Oxford University Press.

Small, C. (1998). *Musicking: The meanings of performing and listening*. University Press of New England.

Smith, B. (1988). Gestalt theory: An essay in philosophy. In B. Smith (Ed.), *Foundations of Gestalt theory* (pp. 11–81). Philosophia.

Snyder, B. (2000). *Music and memory. An introduction*. MIT Press.

Spence, C. (2011). Crossmodal correspondences: A tutorial review. *Attention, Perception, & Psychophysics, 73*(4), 971–995.

Spence, C. (2015). Cross-modal perceptual organization. In J. Wagemans (Ed.), *The Oxford handbook of perceptual organization* (pp. 639–654). Oxford University Press.

Spence, C. (2019). On the relative nature of (pitch-based) crossmodal correspondences. *Multisensory Research, 32*(3), 235–265. https://doi.org/https://doi.org/10.1163/22134808-20191407

Spence, C., & Di Stefano, N. (2022). Crossmodal harmony: Looking for the meaning of harmony beyond hearing. *i-Perception, 13*(1). https://doi.org/10.1177/20416695211073817

Spitzer, M. (2003). *Metaphor and musical thought*. University of Chicago Press.

Stadler, M. (1985). Das Schicksal der nichtemigrierten Gestaltpsychologen im Nationalsozialismus. In C. F. Graumann (Ed.), *Psychologie im Nationalsozialismus* (pp. 139–164). Springer. https://doi.org/10.1007/978-3-642-70064-4_6

Stephan, A. (2006). Emergence. In L. Nadel (Ed.), *Encyclopedia of cognitive science*. John Wiley & Sons. https://doi.org/https://doi.org/10.1002/0470018860.s00112

Stevens, D. (2016). Symphonic hearing: Mastering harmonic dictation using the Do/Ti test. *Journal of Music Theory Pedagogy, 30*, 111–176. https://jmtp.appstate.edu/symphonic-hearing-mastering-harmonic-dictation-using-doti-test

Stevens, D. (2020). In search of hidden treasures: An exercise in symphonic hearing. In *The Routledge companion to music theory pedagogy* (pp. 211–216). Routledge.

Stiles, N. R. B., & Shimojo, S. (2015). Sensory substitution: A new perceptual experience. In J. Wagemans (Ed.), *The Oxford handbook of perceptual organization* (pp. 655–671). Oxford University Press. https://doi.org/10.1093/oxfordhb/9780199686858.013.050

Stillie, B., & Moir, Z. (2021). A good pair of ears: Conceiving of and developing aural skills in popular music education. In K. D. Cleland & P. Fleet (Eds.), *The Routledge companion to aural skills pedagogy* (pp. 179–189). Routledge.

Stumpf, C. (1883). *Tonpsychologie, Vol. I*. S. Hirzel.

Stumpf, C. (1890). *Tonpsychologie, Vol. II*. S. Hirzel.

Swanwick, K. (1994). *Musical knowledge: Intuition, analysis and music education*. Routledge.

Tagg, P. (2004). Gestural interconversion and connotative precision. *Film International, 13*(1), 20–31.

Tagg, P. (2013). *Music's meanings: A modern musicology for non-musos*. Mass Media Music Scholars' Press.

Tagg, P., & Clarida, B. (2003). *Ten little title tunes: Towards a musicology of the mass media*. Mass Media Music Scholars' Press.

Tan, D. (2015). Beyond energetics: Gestalt psychology in Ernst Kurth's Musikpsychologie. *Theoria, 22*, 99–130.

Tenney, J. (2000). *Meta + Hodos: A phenomenology of 20th-century musical materials and an approach to the study of form and Meta Meta + Hodos* (Second ed.). Frog Peak Music.

Thackray, R. (1993). *Aural awakening: A course of aural training and general musicianship for students and teachers* (Revised ed.; orig. pub. 1978). CIRCME.

Thalmann, M., Souza, A. S., & Oberauer, K. (2019). How does chunking help working memory? *Journal of Experimental Psychology: Learning, Memory, and Cognition, 45*(1), 37–55.

Thoresen, L. (2015). *Emergent musical forms: Aural explorations*. University of Western Ontario.

Trubitt, A. R., & Hines, R. S. (1979). *Ear training and sight-singing: An integrated approach*. Schirmer Books.

Tye, M. (2021). Qualia. *The Stanford encyclopedia of philosophy (fall 2021 edition)*. https://plato.stanford.edu/archives/fall2021/entries/qualia/

VanHandel, L. (2020). Music theory and working memory. In L. VanHandel (Ed.), *The Routledge companion to music theory pedagogy* (pp. 3–12). Routledge.

Wagemans, J. (2015). Historical and conceptual background: Gestalt theory. In J. Wagemans (Ed.), *The Oxford handbook of perceptual organization* (pp. 3–20). Oxford University Press.

Wagemans, J., Elder, J. H., Kubovy, M., Palmer, S. E., Peterson, M. A., Singh, M., & von der Heydt, R. (2012). A century of Gestalt psychology in visual perception: I. Perceptual grouping and figure-ground organization. *Psychological bulletin, 138*(6), 1172–1217. https://doi.org/10.1037/a0029333

Walther-Hansen, M. (2020). *Making sense of recordings: How cognitive processing of recorded sound works*. Oxford University Press.

Wertheimer, M. (1912). Experimentelle Studien über das Sehen von Bewegung. *Zeitschrift für Psychologie, 61*, 161–265.

Wertheimer, M. (1923/1938). Laws of organization in perceptual forms. In W. D. Ellis (Ed.), *A source book of Gestalt psychology* (pp. 71–88). Kegan Paul, Trench, Trubner & Company. https://doi.org/10.1037/11496-005

Wertheimer, M. (1924/1938). Gestalt theory. In W. D. Ellis (Ed.), *A source book of Gestalt psychology* (pp. 1–11). Kegan Paul, Trench, Trubner & Company. https://doi.org/10.1037/11496-001

Williams, C. (2019). *The Nashville number system* (7th ed.). Chas Williams.

Winawer, J., Witthoft, N., Frank, M. C., Wu, L., Wade, A. R., & Boroditsky, L. (2007). Russian blues reveal effects of language on color discrimination. *Proceedings of the National Academy of Sciences, 104*(19), 7780–7785. https://doi.org/doi:10.1073/pnas.0701644104

Windsor, W. L. (2004). An ecological approach to semiotics. *Journal for the Theory of Social Behaviour, 34*(2), 179–198.

Windsor, W. L., & de Bézenac, C. (2012). Music and affordances. *Musicae Scientiae, 16*(1), 102–120. https://doi.org/10.1177/1029864911435734

Woods, A. T., Spence, C., Butcher, N., & Deroy, O. (2013). Fast lemons and sour boulders: Testing crossmodal correspondences using an internet-based testing methodology. *i-Perception, 4*(6), 365–379. https://doi.org/10.1068/i0586

Woods, K. J., Siegel, M. H., Traer, J., & McDermott, J. H. (2017). Headphone screening to facilitate web-based auditory experiments. *Attention, Perception, & Psychophysics, 79*(7), 2064–2072.

Woody, R. H. (2020). Musicians' use of harmonic cognitive strategies when playing by ear. *Psychology of Music, 48*(5), 674–692. https://doi.org/10.1177/0305735618816365

Wundt, W. M. (1874). *Grundzüge der physiologischen Psychologie*. Wilhelm Engelmann.

Wöllner, C., (Ed.). (2017). *Body, sound and space in music and beyond: Multimodal explorations*. Taylor & Francis.

Zbikowski, L. M. (2002). *Conceptualizing music: Cognitive structure, theory, and analysis*. Oxford University Press.

Zbikowski, L. M. (2017a). *Foundations of musical grammar*. Oxford University Press.

Zbikowski, L. M. (2017b). Music, analogy, and metaphor. In *The Routledge companion to music cognition* (pp. 501–512). Routledge.

Zouhar, V. (2013). Bohuslav Martinů's notes on Janáček's introduction to the "Moravian Folksongs Newly Collected" (Národní písně moravské v nově nasbírané). *Musicologica Brunensia, 48*(2), 191–199.

APPENDICES

Appendix 1: Digital consent form for survey

Appendix 2: Harmonic luminosity stimuli

Appendix 1

Consent

This survey is part of a PhD research project at the Centre for Educational Research in Music (CERM) at the Norwegian Academy of Music in Oslo, and is conducted by Ville Langfeldt. The goal of the project is to examine aspects of musical harmony perception in musicians and non-musicians.

The experiment is made using Gorilla, a behavioural science research tool which allows the recording of responses and reaction time data.

By taking part in this experiment, metrics will be recorded and uploaded to Gorilla's Servers. This includes background information and response choices during the experiment. No personal or identifying information (including IP addresses) will be gathered or stored. All participation in this experiment is anonymous. We are not able to identify individual participants.

You have the right to withdraw from the experiment at any time—simply close the browser containing this experiment. If you leave the experiment at any time before the end, this will be taken as 'withdrawal' from the experiment and your metrics data will be deleted.

Thank you for agreeing to take part in this experiment! Before we continue, we need your consent to the following:

1. I consent to performing the task online.
2. I understand that my responses are being recorded and stored securely in a database, and I give my consent to this.
3. I understand that my responses will be used anonymously in the aforementioned research project. I also understand that my responses may be used anonymously for secondary research in the future.

☐ **I consent to this**

Next

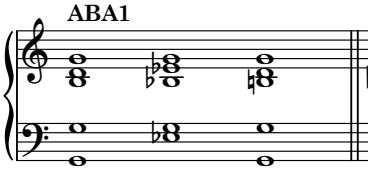
Appendix 2

Appendix 2: Harmonic luminosity stimuli

Numbers in parentheses refer to percentage of hypothesis-supporting answers in educated musician / amateur musician / non-musician populations respectively, plus total percentage across all groups (EM / AM / NM / Total)

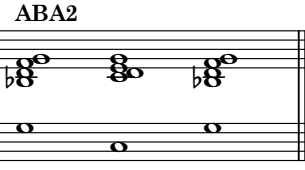
Session 1

ABA1



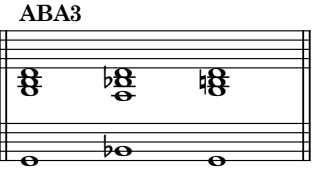
Hypothesis: darkening
(41.9 / 55.0 / 45.8 / 44.9)

ABA2



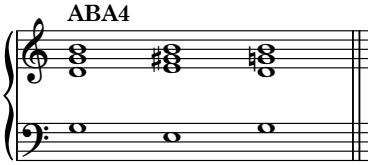
Hypothesis: brightening
(67.6 / 67.5 / 22.9 / 58.5)

ABA3



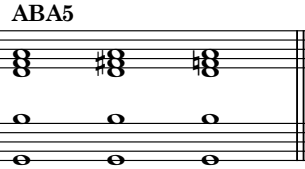
Hypothesis: darkening
(27.0 / 45.0 / 47.9 / 34.3)

ABA4



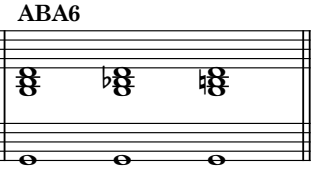
Hypothesis: brightening
(77.0 / 75.0 / 68.8 / 75.0)

ABA5



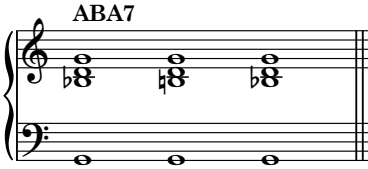
Hypothesis: brightening
(63.5 / 32.5 / 35.4 / 52.5)

ABA6



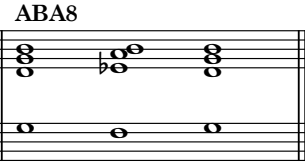
Hypothesis: darkening
(95.9 / 90.0 / 66.7 / 89.0)

ABA7



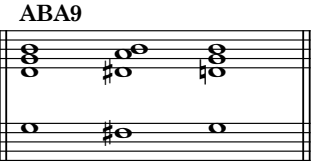
Hypothesis: brightening
(84.5 / 60.0 / 35.4 / 70.3)

ABA8



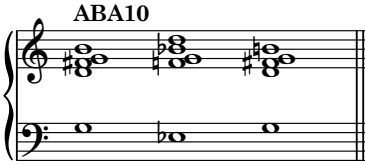
Hypothesis: darkening
(70.9 / 85.0 / 41.7 / 67.4)

ABA9



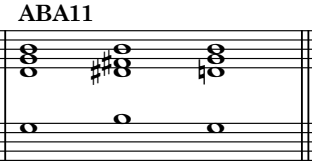
Hypothesis: brightening
(36.5 / 42.5 / 39.6 / 38.1)

ABA10



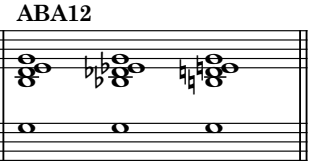
Hypothesis: darkening
(28.4 / 20.0 / 10.4 / 23.3)

ABA11



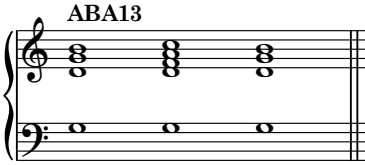
Hypothesis: brightening
(59.5 / 72.5 / 72.9 / 64.4)

ABA12



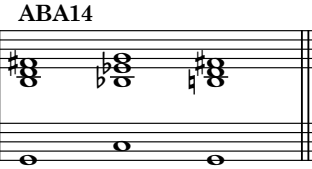
Hypothesis: darkening
(71.6 / 70.0 / 64.6 / 69.9)

ABA13



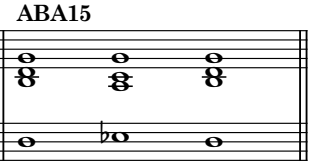
Hypothesis: darkening
(9.5 / 22.5 / 27.5 / 15.3)

ABA14



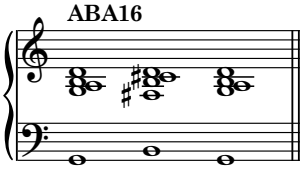
Hypothesis: darkening
(43.9 / 30.0 / 25.0 / 37.7)

ABA15



Hypothesis: darkening
(73.6 / 85.0 / 64.6 / 73.7)

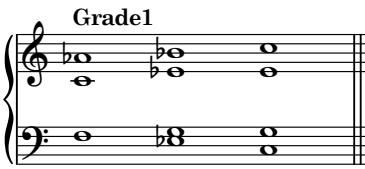
ABA16



Hypothesis: brightening
(15.5 / 27.5 / 20.8 / 18.6)

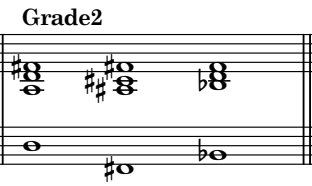
Session 2

Grade1



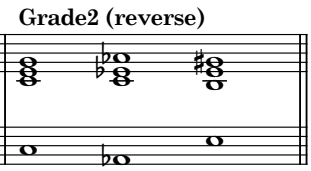
Hypothesis: brightening
(90.5 / 85.0 / 87.5 / 89.0)

Grade2

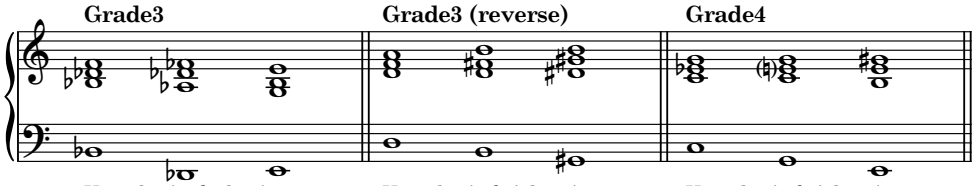


Hypothesis: brightening
(32.4 / 30.0 / 25.0 / 30.5)

Grade2 (reverse)



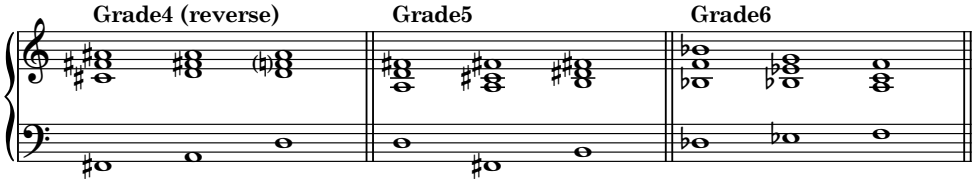
Hypothesis: darkening
(4.7 / 12.5 / 10.4 / 7.2)



Hypothesis: darkening
(65.5 / 87.5 / 75.0 / 71.2)

Hypothesis: brightening
(10.8 / 12.5 / 33.3 / 15.7)

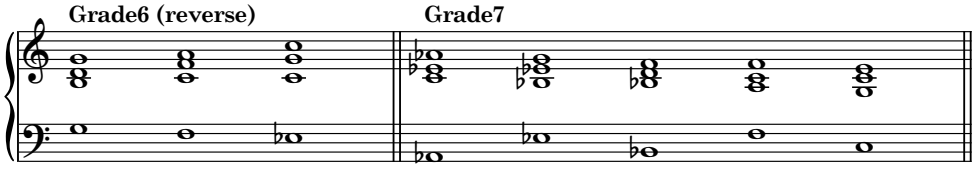
Hypothesis: brightening
(86.5 / 75.0 / 54.2 / 78.0)



Hypothesis: darkening
(69.6 / 72.5 / 39.6 / 64.0)

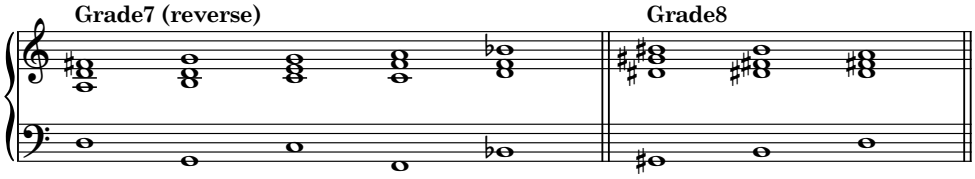
Hypothesis: brightening
(65.5 / 57.5 / 31.3 / 57.2)

Hypothesis: brightening
(86.5 / 67.5 / 47.9 / 75.4)



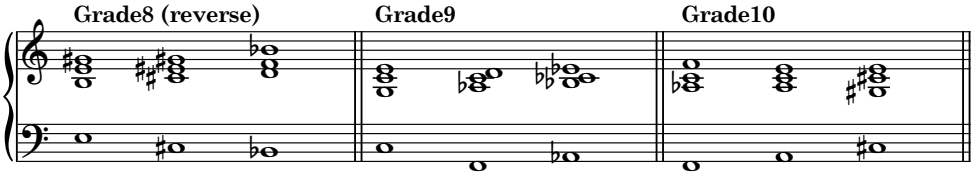
Hypothesis: darkening
(64.9 / 37.5 / 10.4 / 49.2)

Hypothesis: brightening
(43.9 / 32.5 / 29.2 / 39.0)



Hypothesis: darkening
(3.4 / 0.0 / 12.5 / 4.7)

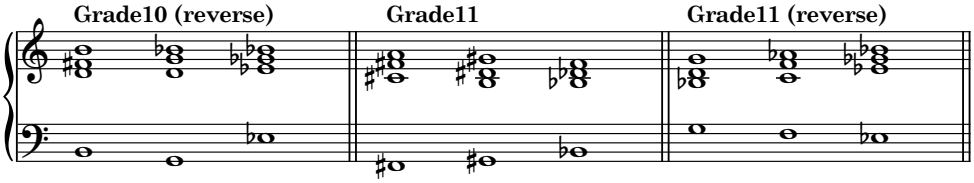
Hypothesis: darkening
(18.9 / 42.5 / 41.7 / 27.5)



Hypothesis: brightening
(64.9 / 65.0 / 60.4 / 64.0)

Hypothesis: darkening
(97.3 / 92.5 / 89.6 / 94.9)

Hypothesis: brightening
(12.8 / 10.0 / 16.7 / 13.1)



Hypothesis: darkening
(52.0 / 60.0 / 27.1 / 48.3)

Hypothesis: brightening
(16.2 / 27.5 / 25.0 / 19.9)

Hypothesis: darkening
(41.2 / 30.0 / 18.8 / 34.7)

What is holistic harmony identification? According to aural training literature, it is how *experts* hear harmony. Unfortunately, this skill cannot be taught—one can only hope to develop it indirectly, after months or years of repeatedly recognizing and labeling particular chords. In his dissertation, Ville Langfeldt challenges this notion by exploring the concept from a range of positions.

The dissertation starts by seeking a clear definition that delineates holistic harmony identification from other approaches to harmonic listening. Through a combination of Gestalt theory and an ecological approach to perception, it continues by examining the perceptual basis of holistic harmony identification, and how it might be experienced from a listener's perspective. Part I of the dissertation ends with a discussion of possible rehearsal strategies for the classroom, based on what Langfeldt calls a "metaphorical" listening approach.

Part II further explores "metaphorical" listening by analyzing conceptual metaphors found in textbooks about harmony. The aim is to shed light on how musical harmony *as sound* is conceptualized through language, and to uncover commonalities in the metaphors we use to describe it. Part III investigates one of these conceptual metaphors specifically: *harmonic luminosity*, or the idea that harmony can express "brightness" and "darkness." In a quantitative study with 236 participants, musicians demonstrate a significantly higher ability than non-musicians to distinguish between "brightening" and "darkening" harmony.

The dissertation's main contribution is the elucidation of the concept of holistic harmony identification, and the proposal for how it may be targeted in the aural training classroom. The dissertation also offers novel perspectives on the role of metaphor in harmonic listening.

Ville Langfeldt (b. 1983) is a music theorist, teacher, and composer.

Find other publications and download them
https://nmh.no/en/research/publications

Norwegian Academy of Music
NMH Publications 2022:7

9 788278 533123