Author Query Form

Journal code: MSR

Transmittal No.: MSR1500

Dear Author,

During the preparation of your manuscript for typesetting, some questions have arisen. They are listed below. Please check your proof carefully and mark all corrections at the appropriate place in the proof (e.g., by using on-screen annotation in the PDF file) or compile them in a separate list. Note: if you opt to annotate the file with software other than Adobe Reader then please also highlight the appropriate place in the PDF file.

Any queries or remarks that have arisen during the processing of your manuscript are listed below and highlighted by flags in the proof.

Location in article	**Query / Remark: Click on the Q link to find the query's location in text** **Please insert your reply or correction at the corresponding line in the proof**
	No queries

Many thanks for your assistance.

BRILL

Multisensory Research 0 (2020) 1–32

MULTISENSORY RESEARCH

brill.com/msr

Crossmodal Correspondence Between Tonal Hierarchy and Visual Brightness: Associating Syntactic Structure and Perceptual Dimensions Across Modalities

Neta Maimon [1,*], **Dominique Lamy** [1,2] **and Zohar Eitan** [3]

[1] The School of Psychological Sciences, Tel Aviv University, POB 39040, Tel Aviv 69978, Israel

[2] The Sagol School of Neuroscience, Tel Aviv University, POB 39040, Tel Aviv 69978, Israel

[3] Buchman-Mehta School of Music, Tel Aviv University, POB 39040, Tel Aviv 69978, Israel

Received 21 July 2019; accepted 2 March 2020

Abstract

Crossmodal correspondences (CMC) systematically associate perceptual dimensions in different sensory modalities (e.g., auditory pitch and visual brightness), and affect perception, cognition, and action. While previous work typically investigated associations between basic perceptual dimensions, here we present a new type of CMC, involving a high-level, quasi-syntactic schema: music tonality. Tonality governs most Western music and regulates stability and tension in melodic and harmonic progressions. Musicians have long associated tonal stability with non-auditory domains, yet such correspondences have hardly been investigated empirically. Here, we investigated CMC between tonal stability and visual brightness, in musicians and in non-musicians, using explicit and implicit measures. On the explicit test, participants heard a tonality-establishing context followed by a probe tone, and matched each probe to one of several circles, varying in brightness. On the implicit test, we applied the Implicit Association Test to auditory (tonally stable or unstable sequences) and visual (bright or dark circles) stimuli. The findings indicate that tonal stability is associated with visual brightness both explicitly and implicitly. They further suggest that this correspondence depends only partially on conceptual musical knowledge, as it also operates through fast, unintentional, and arguably automatic processes in musicians and non-musicians alike. By showing that abstract musical structure can establish concrete connotations to a non-auditory perceptual domain, our results open a hitherto unexplored avenue for research, associating syntactical structure with connotative meaning.

Keywords

Crossmodal correspondences, tonality, musical syntax, visual brightness, Implicit Association Test

---

[*] To whom correspondence should be addressed. E-mail: netacello@gmail.com

DOI:10.1163/22134808-bja10006

1. Introduction

A vast array of empirical studies, using converging psychophysical, cognitive and neuropsychological paradigms, have demonstrated that humans systematically associate “seemingly unrelated features from different sensory modalities” (Parise, 2016), such as auditory pitch and visual brightness, or loudness and visual size (for research reviews, see Marks, 2004; Spence, 2011; Walker, 2016a). Such crossmodal correspondences (CMC) may affect a range of basic perceptual and information-processing domains and processes. These include crossmodal binding in time and space (Parise and Spence, 2009), perceptual learning (Brunel *et al.*, 2015), selective attention (Marks, 2004), and even the perception of basic sensory dimensions, such as spatial location or movement direction (Maeda *et al.*, 2004; Pratt, 1930).

Crossmodal correspondences are particularly ubiquitous in musical contexts, pervading musical vocabulary (e.g., ‘high pitch’, ‘bright sound’), music notation, performance gestures such as conductors’ movements (Globerson *et al.*, in press), and musical patterns conventionally associated with image or text, such as 16th century ‘madrigalisms’ (Wilson *et al.*, 2001) or ‘Mickey-Mousing’ music in 20th century cartoons (Prendergast, 1992). Recent empirical music cognition research has demonstrated that CMC affect various aspects of music-related perception, cognition and action. For instance, music-induced visual imagery (Eitan and Granot, 2006) as well as music-related motion (Kohn and Eitan, 2016; Naveda and Leman, 2010) systematically reflect directed changes in auditory parameters such as pitch, loudness and tempo. In both visual imagery and bodily motion, for instance, listeners associate pitch ‘ascent’ and ‘descent’ with bodily rise and fall, respectively (for additional examples of recent music-related CMC research see Eitan, 2017).

Like most research of crossmodal correspondences, studies of CMC in musical contexts focus mainly on basic perceptual dimensions, which are shared by most auditory domains, like pitch and loudness. Little CMC research, however, examines whether intrinsically musical features such as harmonic and melodic pitch intervals, or higher-level musical structures, such as tonal or metric hierarchy, evoke crossmodal correspondences. Are metrical downbeats, for instance, perceived as ‘heavier’ and ‘lower’ than upbeats? Or, are chromatic (out of key) tones associated with darker visual stimuli than tonally stable tones?

Such investigations may have important implications for music cognition research and for the study of crossmodal correspondences in general. With respect to music cognition, they test the intriguing hypothesis that fundamental cognitive schemas that underlie musical structure may also underlie — at least in part — music’s associations with non-auditory domains. For CMC research, these investigations have the potential to expand a field currently focusing on associations of basic sensory dimensions, by examining whether and how higher-level cognitive schemata may partake in generating crossmodal mappings. Specifically, the present investigation may suggest how syntactical function may imply specific crossmodal connotations. In this paper we focus on tonality, the pitch structure underlying much of Western music. Tonality is, in important ways, a musical syntax: a set of rules and practices, implicitly understood by listeners, governing closure and continuity, stability and tension, in melodic and harmonic sequences (Koelsch, 2011; Lerdahl and Jackendoff, 1983). Hence, the investigation of crossmodal connotations of tonal structures may suggest an intriguing path, relating syntactical structure and connotative meaning, and opens the way for the study of comparable relationships in non-musical domains, such as language syntax. The present paper, the first in a series of ongoing studies (Maimon *et al.*, 2017, 2018, 2019), introduces such research through two experiments, in which explicit and implicit associations of tonal stability and visual brightness are examined.

1.1. Tonal Stability

In music theory, the term 'tonality' denotes a system organizing pitch relationships in a musical work, both melodically (i.e., with regard to pitch succession) and harmonically (with regard to pitch simultaneities, or 'chords,' and their succession), in a hierarchy of stability and closure. Importantly, for each tonal context, or musical key, pitches are organized with reference to one maximally stable pitch class (see Note 1), the *tonic* or *tonal center*.

What is 'tonal stability'? Stable pitches and chords are associated with points of closure (endings of segments) in tonal music — an association underlying listeners' perception of musical closure (Bigand *et al.*, 2006; Boltz, 1989). Furthermore, unstable tones tend to proceed to the nearest stable tones (but not vice versa), a distributional fact affecting listeners' expectations (Huron, 2006). Thus, unstable tones or chords tend to suggest to listeners a sense of tension and continuity, while the stable tones/chords following them suggest resolution of that tension (Lerdahl and Krumhansl, 2007). Tonal stability also correlates with pitches' occurrence frequency: within a tonal context, stable pitches are more frequent than instable ones. Compatibly, listeners tend to perceive stable pitches as better fitting a primed tonal context than unstable ones (Krumhansl, 1990).

In Western tonal music, tonal stability is governed by musical keys. A musical key is defined by its tonic and its mode (major or minor). The tonic — which may be any of the 12 pitch classes available in Western music — is, as noted above, the maximally stable, most closural note in a given key. For each tonic, two different modes are available, major or minor, each established by a different pattern of pitch intervals between the tonic and the other constituent

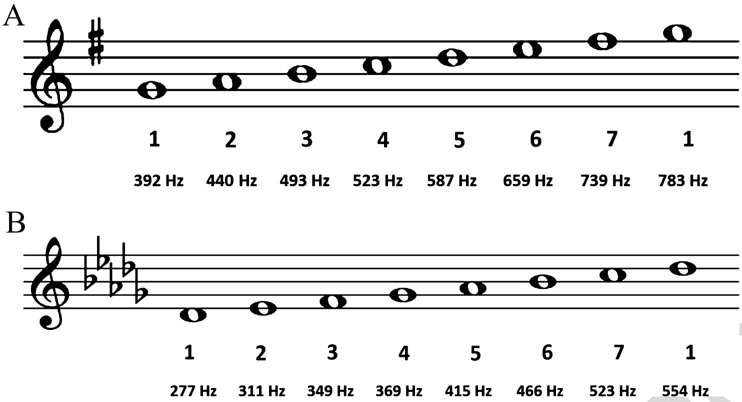

Figure 1. Musical notation of the two major scales used in Experiments 1A and 1B, G major (A), and D-flat major (B). Arabic numerals below the notes denote their scale degree identity (1–7). Fundamental frequencies (F0, in Hz) are marked below the scale degree numbers (note, however, that each scale degree can be represented by different pitch classes — notes sharing the same name, situated one or several octaves apart — e.g., A1; 55 Hz; A2, 110 Hz; A3, 220 Hz; A4, 440 Hz, etc).

tones. Thus, the keys of C major and C minor share the same tonic — the pitch class C — but have different modes; that is, the relationships of some of their other constituent pitches to the tonic differs.

Each major or minor key chiefly utilizes seven out of the 12 available pitch classes, its diatonic scale degrees. Diatonic scale degrees are commonly represented as a musical scale, arranged in order of pitch height, with the tonic note presented as first (lowest) note, or scale degree 1, and the other diatonic degrees presented and numbered accordingly (2–7; see Fig. 1A, B, for scalar representations in music notation of the two keys used in our experiments — G major and D-flat major). The most stable diatonic notes are the tonic (scale degree 1) and the other constituents of the tonic triad (the chord associated with the tonic note) — scale degrees 3 and 5. Other diatonic degrees (2, 4, 6, 7) are relatively instable, evoking tension that may be resolved when their more stable neighbors (1, 3, or 5 scale degrees) follow (Lerdahl and Krumhansl, 2007).

While simple tonal pieces (e.g., many western European folk songs) may exclusively use the seven diatonic scale degrees, more complex tonal music also applies the remaining five pitch classes (termed ‘chromatic tones’ by musicians), which are conceived (by music theorists) and perceived (by listeners; see Krumhansl, 1990, for review of empirical studies) as the least stable tones in a tonal context. These ‘out of key’ notes tend to evoke strong tension, resolved when their nearest diatonic (within-key) notes follow. In sum, then, musical theory, as well as music cognition research, distinguishes three main levels of tonal stability: tonal triad members (scale degrees 1, 3, 5) are the most stable; other diatonic notes (scale degrees 2, 4, 6, 7) are less stable, and tend to resolve to nearby stable scale degrees; and the remaining chromatic (out of key) notes are the least stable, strongly implying resolution by their diatonic neighbors. Figure 2A demonstrates this hierarchy.

Importantly, the hierarchy of tonal stability is always relative to a key's tonic, such that the very same pitch may be highly stable in one key, and highly unstable in another. For instance, in the key of G major, the pitch class C-sharp/D-flat (which is a chromatic, 'out of key' note) would be extremely unstable, while G, the tonic, would be the maximally stable pitch class. In contrast, in the key of D-flat major, D-flat (the tonic note of that key) would be maximally stable, while G (here, a chromatic note) would be highly unstable. Figures 2A–C demonstrate this relativity: Fig. 2A is a general representation of the tonal hierarchy described above. Figures 2B and C demonstrate the same hierarchy as applied to two different keys — G major and D-flat major. Note that while the tonal schema remains the same in both keys, the representatives (i.e., specific pitch classes) of its 'slots' (the scale degrees) radically change, with stable degrees in one key becoming instable in the other.

Tonality underlies the bulk of Western music ('classical' as well as popular) since the 17th century and has been described and modeled in detail by music theorists (e.g., Fétis, 1844; Rameau, 1722; Schoenberg, 1978). In recent decades, the psychological reality of tonality as a cognitive schema orienting the listener has been strongly established empirically (see Krumhansl, 2004 and Shanahan, 2017 for research surveys). Studies applying converging experimental paradigms — explicit measurements, such as sung continuations and goodness-of-fit ratings (e.g., Carlsen, 1981; Cuddy and Lunney, 1995; Krumhansl, 1990), as well as implicit ones, such as musical priming and event-related potentials (ERP; e.g., Granot and Donchin, 2002; Marmel and Tillman, 2009) — have suggested that listeners implicitly abstract a tonal hierarchy, and the sets of melodic and harmonic expectancies it entails, closely matching those conjectured by music theorists' models.

Importantly, in addition to tonality's 'syntactic' roles, musicians' accounts suggest that tonal hierarchy may also evoke expressive and crossmodal connotations: different scale degrees may 'feel' differently, and associate in consistent ways with diverse non-auditory domains (see Rothfarb, 2001, for a historical review). Accordingly, tonal relationships were mapped onto basic natural forces like gravity (Rameau, 1722) or magnetism (Arnheim, 1984; Larson, 2004), spatial image schemas like center/periphery (e.g., tonal center, outside the key; Larson and VanHandel, 2005), top/bottom, or front/back (Spitzer, 2003), and goal-oriented motion (Schenker, 1906). Often, such relationships were presumed to express or evoke aspects of human emotion, such as passionate desire toward the tonic (Rameau, 1722), or violent emotions

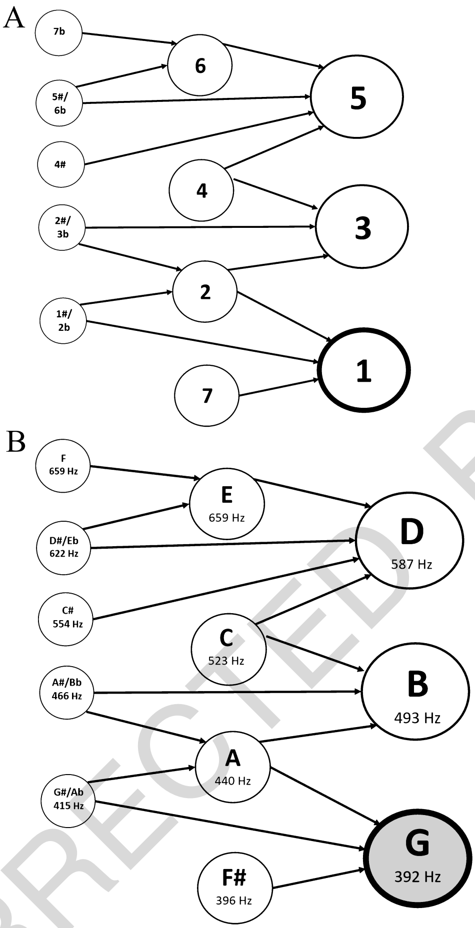

Figure 2. Visual representations of tonal hierarchy. Figure 2A presents generalized scale degree representation, while panels B and C, respectively, present these relationships in G major and D-flat major keys, replacing the scale degree numbers in A with the pitch classes representing these scale degrees in each key (fundamental frequencies are added below pitch class names, corresponding to those of pitches in Fig. 1). The larger and bolder scale degrees on the right are the most stable (stable diatonic), the middle layer presents other diatonic scale degrees — less stable yet belonging to the respective key (unstable diatonic), and the third (leftmost) layer presents the least stable tones (chromatic, ‘out of key’ tones). Arrows between scale degrees represent the typical motion direction between them — from an instable degree to a stable one. Note that while the tonal schema represented in 2A remains the same in both keys (2B, 2C), the representatives (i.e,. specific pitch classes) of its ‘slots’ (the scale degrees) radically change, with stable degrees in one key becoming instable in the other. For instance, the pitch class G (gray circles), which is the tonic note — the most stable note — in G major (Fig. 2b) is an instable chromatic note in D-flat major (Fig. 2c).

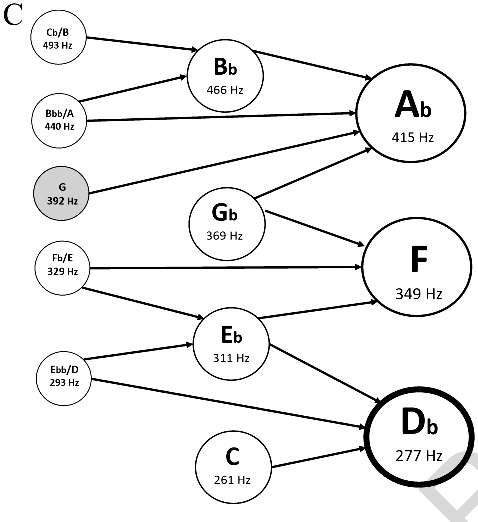

Figure 2. (Continued.)

depicted by chromaticism (Fétis, 1844). Indeed, crossmodal and emotional metaphors for tonality are often strongly entangled, as in the common dichotomy of 'dark' chromaticism vs 'bright' diatonicism, which underlies both crossmodal and emotional musical connotations in music ranging from Gesualdo to Wagner (Boulez, 1986, p. 254).

As prevalent as such intuitive mappings are, only a few studies have put them to empirical test. Moreover, these studies have used only explicit measures relying on conscious introspection and on verbal responses, such as free descriptions or adjective ratings (Arthur, 2018; Huron, 2006). Their results suggest that people indeed find it more appropriate to associate a particular scale degree with one metaphorical description (e.g., 'dark') than with its antonym ('bright'). They do not, however, examine whether such associations, like CMC between basic perceptual dimensions (see above), arise when participants are not required to make explicit introspective judgments, but can rather be reflected implicitly through behavior (e.g., reaction times), perception (e.g., modulating perceived judgment of concurrent visual dimensions), or neural activity (e.g., EEG measures). Addressing these pending questions is important, because responses in previous experiments could mainly reflect explicit academic learning of conventional metaphors, which clearly indicate to participants what the 'correct' answers are — particularly considering that most participants were trained musicians.

In the present research, we attempt to address the above issues by (a) using both explicit and implicit experimental measures, (b) lessening the role of verbal responses, and (c) systematically comparing responses of trained musicians with those of musically untrained participants, who lack the conceptual baggage which may underlie musicians' explicit responses. In this paper, we investigate CMC of tonal stability with a single perceptual dimension, visual brightness. The methods employed here, however, may be applied to studying the associations of tonality with other non-auditory dimensions, such as visual size or spatial position — a line of research currently conducted in our lab (see Maimon *et al.*, 2017, 2018 for preliminary reports).

We chose to focus on visual brightness for several reasons. First, this dimension is preeminent in CMC research, which has established its correspondences with basic auditory dimensions, including pitch height, loudness, and timbre (see Eitan, 2013; Marks, 2004; Spence, 2011, for research reviews). Second, it has been associated with tonal stability in musicians' conventions and subjective accounts (e.g., 'dark chromaticism') as well as in the few existing relevant behavioral studies (Arthur, 2018; Huron, 2006). Finally, brightness associates cognitively and perceptually with affective dimensions, particularly valence, as both conventional metaphors (e.g., dark/bright mood) and experimental work (e.g., Meier *et al.*, 2004, 2007) suggest. Possible interactions between emotional and crossmodal connotations of brightness in its mapping onto tonal stability further enhance the interest in the brightness–tonality CMC.

To investigate this association, we used two experimental paradigms, each tapping a different processing level. In Experiments 1A and 1B, we used an explicit test that relied on Krumhansl's probe–tone paradigm (e.g., Krumhansl and Kessler, 1982). In that paradigm, participants hear on each trial a key-establishing 'context element,' (e.g., a cadence — a sequence of chords conventionally associated with tonal closure), serving to prime a specific key, followed by a single tone, the 'probe.' For each context element in a given key, probes consist of the twelve pitch classes, with a different pitch class presented in each trial. Participants are asked to rate how well the probe fits with the context element that preceded it, on a 1–7 Likert scale. In experiments based on this paradigm (reviewed in Krumhansl, 1990), participants' ratings generally reflected tonal hierarchy, as theorized by musicians, with more stable probe tones rated as better fitting the context element than less stable ones: tonic triad members (1, 3, 5) rated higher than other diatonic tones (2, 4, 6, 7), while chromatic ('out of key') tones received the lowest rating, i.e., were perceived as least fitting the key-establishing context.

In the present study (Experiment 1A), participants performed two probe–tone tasks. The first, a goodness-of-fit (GOF) task, replicated the original probe–tone procedure, as described above. The second (the brightness task), used a similar procedure, except that instead of providing a numerical rating to the goodness of the element–probe fit, participants reported, from their subjective viewpoint, how (visually) bright the probe 'felt' in relation to the preceding element. To do that, they selected one of seven circles varying from dark to bright and presented in a row. If tonal stability is explicitly associated with brightness, we expected participants to select brighter circles when the probe belonged more stable tones: hence, brighter circles would be selected more for tonic triad members (1, 3, 5) than for other diatonic scale degrees, while the darkest circles would be selected for chromatic probe tones. We also expected participants' ratings in the two tasks (GOF and brightness) to significantly correlate.

In Experiment 2, we used an implicit test, adapted from Parise and Spence's (2012) crossmodal version of the Implicit Association Test (IAT). In this paradigm, the stimulus set consists of two auditory and two visual stimuli. Two stimuli — one auditory and one visual — are assigned to the same response key in a given block of trials. Only one of the four stimuli is presented on each trial, and participants are required to respond as fast as possible using one of the two possible responses. The objective is to determine whether one visual stimulus is more strongly associated with one of the auditory stimuli (compatible stimulus) than with the other (incompatible stimulus): this can be inferred if participants' performance is better (faster reaction time [RT], fewer errors) when compatible stimuli are assigned to the same response key.

Here, we used a tonally stable and a tonally unstable stimulus for the auditory dimension, and two levels of brightness (a dark circle and a bright circle) for the visual dimension. If tonal stability is implicitly associated with brightness, we expected participants to respond faster when the stable tone and the bright circle were assigned to the one response key and the unstable tone and the dark circle to the other response key, compared to when the alternative stimulus–response assignment was used.

In all experiments, we included both musically-trained and untrained participants in order to examine the role of expertise in processing the CMC.

2. Experiment 1A

2.1. Methods

2.1.1. Sample Size Selection

On the basis of the study by Krumhansl and Kessler (1982), we calculated the sample size required to observe a significant effect of tonal hierarchy. Since they conducted multiple comparisons, we based our power analysis on the comparison between diatonic scale degrees and non-diatonic scale degrees, which had the smallest effect size. We conducted this analysis with G*Power (Erdfelder *et al.*, 1996) using an alpha of 0.05, and a power of 0.80. We found the minimum required sample size to be four participants. As we examined possible interactions of this effect with musical training (musicians vs non-musicians) and task (goodness of fit vs brightness matching), the minimum sample size was 16.

2.1.2. Participants

Forty undergraduate and graduate students from Tel Aviv University, 20 musicians (3 females, mean age = 28.65, SD = 6.52) and 20 non-musicians (12 females, mean age = 29.21, SD = 4.67), served as participants. The musicians had an average of 18.4 (SD = 6.56) years of musical experience (with a minimum of nine years) and an average of 10.75 (SD = 6.94) years of music theory studies (with a minimum of five years). They all currently played and performed music. They were paid 20$ per hour for their participation. The non-musicians had an average of 1.42 (SD = 1.42) years of musical experience (with a maximum of three years in childhood), no music theory education, and none of them currently played music. They were given a one-hour experiment credit for their participation.

2.1.3. Apparatus

The stimuli were presented on a computer with an Intel (Santa Clara, CA, USA) Core i7 CPU 920 processor, 2.67 GHz speed, and 2.98 GB RAM memory. Auditory stimuli were delivered through Sennheiser (Wedemark, Germany) 210HD precision headphones. These were connected to the computer by a Terratec (Nettetal, Germany) Producer, Phase 24 sound card. Stimulus loudness was measured through Brüel and Kjær (Nærum, Denmark) type 2232 noise meter, measuring A-weighted dB. The computer screen was a 17-inch Lenovo (Quarry Bay, Hong Kong) LCD, with a screen refresh frequency of 85 Hz and 1024 × 768-pixel resolution. The experiment was programmed and run in Matlab (MatWorks, Natick, MA, USA).

2.1.4. Stimuli

Auditory Stimuli. Pitch height is known to be strongly associated with visual brightness (Melara and Marks, 1990). In order to minimize pitch height effects, all auditory stimuli (both contexts and probes) were created using Shepard tones (Shepard, 1964), Shepard tones are tones with a specific pitch chroma (pitch class), sounded in five octaves simultaneously. Following the standard protocol presented in Shepard, 1964, we created the tones using a loudness envelope across the frequency range of 77.8–2349 Hz, with partials uniformly decreasing in intensity toward low and high frequencies (Shepard, 1964). Shepard tones generate a clear pitch chroma (pitch class) but an ambiguous pitch height (i.e., the register in which a pitch is perceived — which partial in a given tone represents its pitch height for a listener — is ambiguous and determined contextually). All stimuli were sounded at 71 dB.

Each trial consisted of a chord or a chord sequence which established a sense of tonal key — the ‘context element’ — followed by a single tone out of the 12 chromatic tones in western tonality — the 'probe'. There were eight possible elements (see audio examples in Supplementary Audio S1): two element types (a triad, the first, third and fifth degrees played simultaneously, presumably perceived as a tonic chord, and a IV–V–I cadence, a sequence of three chords which imply a strong notion of closure in the tonal system). Each of these two types was played in four keys (G major, G minor, D-flat major and D-flat minor). There were 12 possible probes: the 12 tones of the chromatic scale (12 notes from C to B). On each trial, an element was played for 0.5 s when it was the tonic chord, and for 0.5 s for each chord followed by a silence of 0.25 s when it was a cadence. Then, following a silence of 1 s, a probe tone was played for 0.5 s. Sound Examples 1–4 present four examples of an element followed by a probe. In Sound Example 1, an element (a cadence) in D-flat major is followed by a stable probe (the tonic note). In Sound Example 2 the same element is followed by an instable probe. In Sound Examples 3 and 4 elements in G major are followed by stable and instable probes, respectively.

The experiment consisted of 18 blocks. The first two blocks served as practice. The elements for these blocks were quasi-randomly drawn from the eight possible elements and consisted of either a cadence or a triad, in either G or D-flat (counterbalanced across subjects). They were followed by 16 experimental blocks, two for each of the eight possible elements, randomly mixed. The same element was used throughout any given block of trials. Each block consisted of 14 consecutive trials. The first two trials served as practice, using two randomly selected probes. In the following 12 experimental trials, the 12 chromatic tones, randomly mixed, were used as probe tones.

Visual Stimuli. For the GOF task, a 1–7 Likert-like scale (e.g., seven outline boxes arranged in a row and enclosing the numbers 1 to 7) was presented in either ascending (from left to right) or descending (from right to left) order. The words 'fits poorly' were written to the left of the number 1, and the words 'fits well' were written to the right of the number 7. For the brightness-matching task, seven gray circles were presented horizontally on a black screen, each subtending 1.60 in diameter, with 10 cd/m^2 between them. Brightness levels spanned the maximal range (0–255) and the intermediate levels were determined by the Stevens' power law (Stevens, 1964), with a coefficient of 0.5: each circle was twice as bright as the immediately preceding circle.

Procedure. The experiment was conducted in a dark sound-attenuating room, with no lighting expect that of the computer screen. Participants were seated 50 cm from the computer screen. The experiment was divided into two sessions, the GOF task session and the brightness-matching task session, administered 7 to 14 days apart ($M = 8.62$). Task order was counterbalanced across participants. The trial randomization for each participant was the same for both tasks.

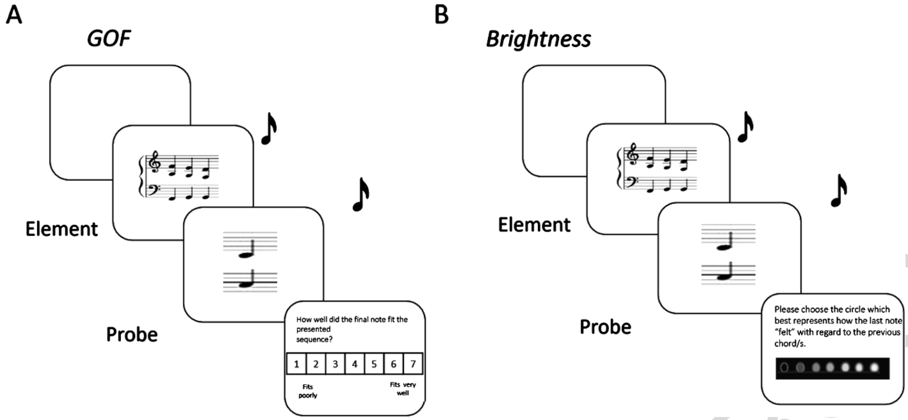

Figure 3. Graphic representation of a trial sequence in the GOF (goodness-of-fit) task (A) and the brightness task (B) of Experiment 1A. In each trial participants heard an auditory context element followed by an auditory probe and were asked to either rate the probe's fitness to the element (GOF task) or choose the appropriate circle (brightness task).

GOF Session. For the trial sequence see Fig. 3A. The 1–7 Likert-like scale appeared on the screen throughout the entire block. After listening to the auditory element followed by the auditory probe, participants were asked to rate how well they felt the probe fit the previous element by clicking the corresponding numeral on the upper row of the keyboard.

Brightness-Matching Session. For the trial sequence see Fig. 3B. The seven brightness circles, presented horizontally on a black screen, ranging in brightness from black to white (either from left to right or from right to left, counterbalanced between participants), appeared on the screen throughout the entire block. After listening to the element followed by the probe, participants were asked to mark the circle which, according to their subjective judgment or 'feeling,' matched each probe tone best. They were required to click the corresponding circle with the cursor.

The experimenter underscored the subjective and intuitive character of the task. She instructed participants to choose whichever strategy they felt was the most suitable to perform the tasks, yet to be consistent and use the same strategy throughout the experiment.

After each block, participants were asked about their confidence in the judgments they had provided. For this measure, we were mainly interested in examining whether task (the musical GOF task versus the crossmodal brightness-matching task) and musical expertise (whether participants had professional musical training) affected how secure participants felt about their intuitions. The text "How confident were you with the ratings of the last block?" appeared on the screen and participants provided a 1–7 rating by pressing the appropriate numeral on the upper row of the keyboard. Then, a 10-s white noise (71 dB) was heard, followed by a count down from 5 to 1. Participants then proceeded to the next block by pressing any key, after a self-pace break.

All written and oral instructions were presented in Hebrew.

2.2. Results

The data from one non-musician was excluded since she did not agree to complete the second session. Preliminary analyses showed no significant effect involving element types (i.e., cadence vs triad) or task order. The data were therefore collapsed across these variables.

A separate repeated-measure analysis of variance (ANOVA) was used for each task (GOF and brightness), with musical training (musician vs non-musician) as a between-participants variable and tonal stability and mode (major/minor) as within-participant variables. The tonal stability variable included three categories: *Stable diatonic* (the tonic chord members — the 1st, 3rd, and 5th degrees of the scale; that is, tones 1, 5, 8 of the 12 chromatic tones in major and 1, 4, 8 in minor), *Unstable diatonic* (2nd, 4th, 6th and 7th degrees of the scale: 3, 6, 10, 12 of the 12 *chromatic tones* in major and 3, 6, 7, 11, 12 in minor, where both options for the 7th scale degree, raised [leading-tone] and natural, were included) and chromatic tones (non-scale degrees: 2, 4, 7, 9, 11 of the 12 tones in major and 2, 5, 7, 10 in minor). For a summary of the results, see Fig. 4.

2.2.1. Goodness-of-Fit Task

The main effect of tonal stability was significant, $F_{2,74} = 147.10$, $p < 0.001$, $\eta_p^2 = 0.799$. Pairwise comparisons using Bonferroni corrections revealed that ratings were higher for the stable diatonic than for the unstable diatonic category, which were higher than ratings for the chromatic-tones category (both ps < 0.001). The main effect of musical training was significant, $F_{1,37} = 15.11$, $p < 0.001$, $\eta_p^2 = 0.29$, with higher ratings for musicians than for non-musicians. The interaction between tonal stability and musical training was significant, $F_{2,74} = 44.12$, $p < 0.001$, $\eta_p^2 = 0.544$. Follow-up comparisons indicated that musicians and non-musicians showed the same significant differences between categories, but musicians were more extreme than non-musicians in their ratings. Generally, these results replicate the findings of earlier probe–tone experiments (e.g., Krumhansl and Kessler, 1982).

There was no significant effect of mode ($F < 1$). However, the interaction between mode and tonal stability was significant, $F_{2,74} = 20.18$, $p < 0.001$, $\eta_p^2 = 0.353$, and was modulated by a higher-order interaction with musical experience, $F_{2,74} = 7.657$, $p = 0.001$, $\eta_p^2 = 0.171$. Follow-up analyses clarified this three-way interaction. They revealed that all the differences between tonal stability categories were significant, in both major and minor modes and

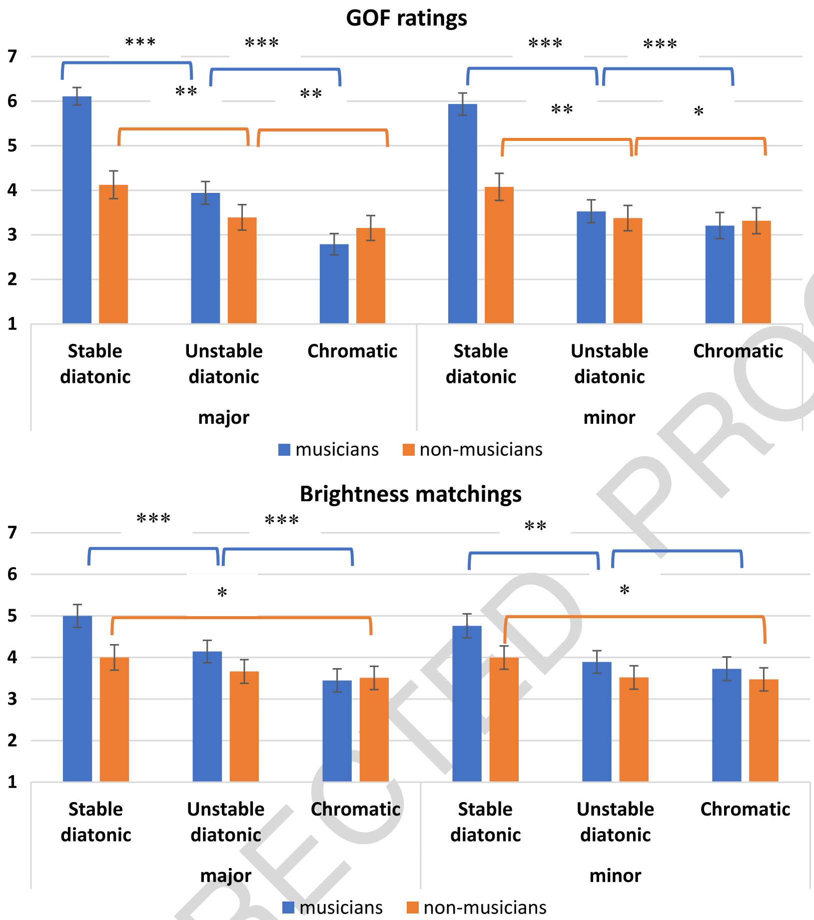

Figure 4. Mean GOF (goodness-of-fit) ratings (upper graph) and brightness matchings (lower graph) in Experiment 1A as a function of tonal stability categories, in major (right panel) and minor (left panel) modes and for musicians (blue bars) and non-musicians (red bars). In GOF ratings 1 = least fitting, 7 = most fitting. In brightness matchings 1 = darkest circle, 7 = brightest circle.

for both musicians and non-musicians (all $ps < 0.05$). However, GOF ratings of tonal stability categories were more extreme within the major mode than within the minor mode, and rating differences between tonal stability categories were more extreme for musicians than for than non-musicians within both modes.

2.2.2. *Brightness-Matching Task*

The main effect of tonal stability was significant, $F_{2,74} = 20.209$, $p < 0.001$, $\eta_p^2 = 0.353$. Pairwise comparisons using Bonferroni corrections revealed that brightness matchings were higher for the stable diatonic than for the unstable diatonic category, and these were higher than brightness matchings for the chromatic-tone category (both ps < 0.001). The main effect of musical training was significant, $F_{1,37} = 8.142$, $p = 0.007$, $\eta_p^2 = 0.18$, with higher overall brightness ratings for musicians than for non-musicians.

Tonal stability interacted both with mode, $F_{2,74} = 5.512$, $p = 0.008$, $\eta_p^2 = 0.13$ and with musical training, $F_{2,74} = 3.34$, $p = 0.041$, $\eta_p^2 = 0.083$. These interactions were modulated by a significant three-way interaction between tonal stability, musical training and mode, $F_{2,74} = 4.568$, $p = 0.017$, $\eta_p^2 = 0.11$. To clarify this interaction, we conducted a separate ANOVA for musicians and for non-musicians, with tonal stability and mode as within-participants variables.

2.2.3. *Musicians*

The main effect of tonal stability was significant, $F(2,38) = 15.8$, $p < 0.001$, $\eta_p^2 = 0.454$. Pairwise comparisons using Bonferroni corrections revealed that brightness matchings were higher for the stable diatonic than for the unstable diatonic category, and these were higher than brightness matchings for the chromatic-tone category (both ps < 0.001). The main effect of mode was not significant, $F < 1$, but interacted with tonal stability, $F_{2,38} = 6.91$, $p = 0.003$, $\eta_p^2 = 0.267$, indicating that the differences in musicians' brightness matchings between tonal stability categories were more extreme within the major than within the minor mode: pairwise comparisons revealed that within the major mode, the brightness-matching differences between tonal stability categories were significant, both ps < 0.01, whereas within the minor mode, the difference between stable- and unstable-diatonic categories was significant ($p = 0.016$), but the difference between the unstable-diatonic and the chromatic-tones categories was not, $p = 0.41$.

2.2.4. *Non-Musicians*

The main effect of tonal stability was significant, $F_{2,36} = 5.186$, $p = 0.01$, $\eta_p^2 = 0.224$. While none of the pairwise comparisons using Bonferroni corrections revealed a difference in brightness matchings between of the three tonal stability categories (both ps > 0.14), there was a significant linear trend, with brightness matchings increasing linearly from chromatic to stable diatonic through unstable diatonic, $F_{1,18} = 5.897$, $p = 0.026$, $\eta_p^2 = 0.274$. There was no other significant effect, all ps > 0.05.

2.2.5. *Correlations Between GOF Ratings and Brightness Matchings*

The results summary is presented in Fig. 5 (Note 2). For each of the 12 scale degrees, mean GOF ratings and brightness-matching scores were calculated separately for the major and minor modes and averaged across participants.

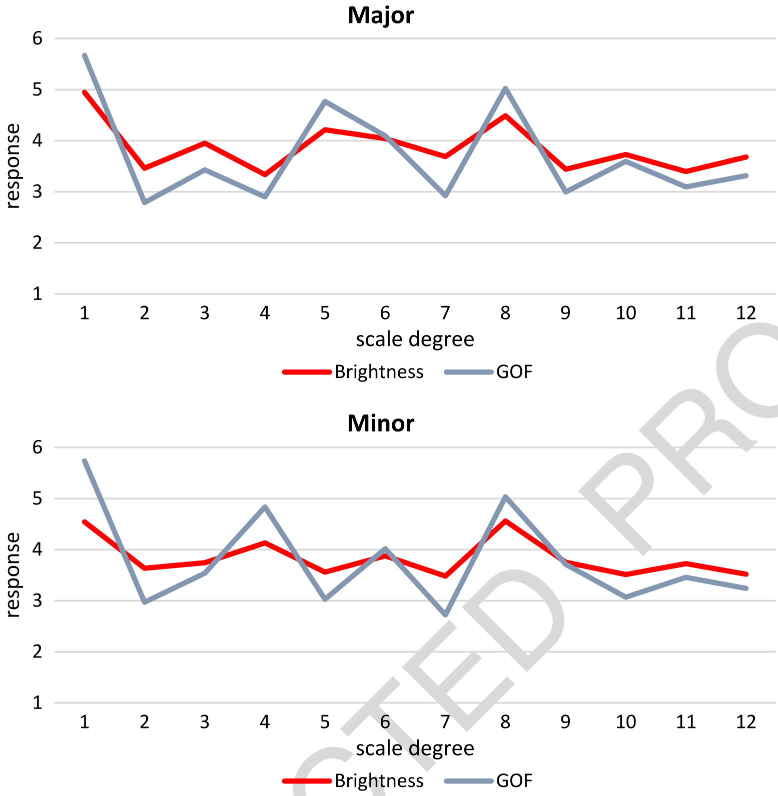

Figure 5. Mean GOF (goodness-of-fit) ratings (blue) and brightness matchings (red) in Experiment 1A as a function of scale degrees, in major (upper graph) and minor (lower graph) modes. In GOF ratings 1 = least fitting, 7 = most fitting. In brightness matchings 1 = darkest circle, 7 = brightest circle.

The results showed highly significant correlations between GOF and brightness ratings, Pearson $r = 0.926$, $p < 0.001$ for major scale degrees and Pearson $r = 0.966$, $p < 0.001$ for minor scale degrees.

In addition, the same correlations were calculated separately for each participant. For non-musicians (in both major and minor modes), 6 out of 19 participants showed a significant positive correlation between GOF ratings and brightness matchings, while none showed a significant negative correlation. For musicians, 10 out of 20 participants showed a significant positive correlation between GOF ratings and brightness matchings in the major mode and 8 out of 20 in the minor mode, while none showed a significant negative correlation in either mode. These correlations were further reinforced by the observation that participants who showed significant correlations within their own GOF ratings and brightness matchings also showed significant correlations with the group's average on both measures, and conversely for participants who did not show a significant correlation between GOF ratings and brightness matching within their own ratings. This notwithstanding, the results do point to individual differences in the extent to which GOF and brightness are correlated, beyond differences related to music training. These differences could be interesting to clarify in further research.

2.2.6. Confidence Ratings

An ANOVA with musical training as a between-participants variable and task as a within-participant variable was conducted on the mean confidence ratings. The main effect of musical training was significant, with musicians being more confident than non-musicians ($M = 5.32$, SD $= 1.35$ vs $M = 4.15$, SD $=$ 1.48), $F_{1,37} = 13.99$, $p = 0.001$. The main effect of task was not significant, $M = 4.8$, SD $= 1.5$ for GOF rating task and $M = 4.67$, SD $= 1.56$ for the brightness-matching task, and neither was the interaction between the two factors, both Fs < 1.

2.3. Discussion

The results of Experiment 1A revealed a clear correspondence between tonal stability and visual brightness. Brighter circles were matched to more stable tonality probes, and brightness matchings and GOF ratings were correlated across the 12 scale degrees. Though this association was significant for both musicians and non-musicians, it was modulated by musical training, which suggests that it relies, at least in part, on conceptual musical knowledge.

One notes that the scalar, linear ordering of the visual stimuli in the brightness task (circles ordered from either bright to dark or dark to bright) might have encouraged participants to construe the task as a ranking task, quite similar to the GOF task, without specifically relying on brightness. A schema relating probe tones and brightness levels along a shared scale could, then, be (at least in part) an artifact of the experiment's design. In Experiment 1B we addressed this issue by replicating the brightness-matching task of Experiment 1A, while ordering the visual stimuli randomly, rather than as a graded brightness scale.

3. Experiment 1B

3.1. Method

3.1.1. Sample Size Selection

On the basis of the main effect of tonal stability in Experiment 1A, we calculated the sample size required to observe a significant effect of tonal hierarchy. We conducted this analysis with G*Power (Erdfelder *et al.*, 1996) using an alpha of 0.05, and a power of 0.80. We again found the minimum required sample size per condition to be four participants. As we examined possible interactions of this effect with musical training (musicians vs non-musicians), the minimum sample size was eight.

3.1.2. Participants

Twenty undergraduate students from Tel Aviv University served as participants, 10 musicians (2 female, mean age = 22.9, SD = 2.84) and 10 non-musicians (5 female, mean age = 24.5, SD = 4.03). The musicians had an average of 14.6 (SD = 3.94) years of musical experience (with a minimum of nine years) and an average of 8.8 (SD = 5.18) years of musical theory studies (with a minimum of five years). They all currently played and performed music. The non-musicians had an average of 1.375 (SD = 0.75) years of musical experience (with a maximum of three years in childhood), no music theory education and none of them currently played music. None of the participants had taken part in Experiment 1A. All participants were paid 10$ per hour for their participation.

3.1.3. Apparatus, Stimuli and Procedure

The apparatus, stimuli and procedure were similar to those of the brightness-matching task of Experiment 1A, except for the following differences. The seven circles were presented in random order rather than in a monotonically increasing or decreasing order. A given random order remained constant in each block and changed between the blocks. Participants were notified that randomization changes would occur during the experiment. Unlike in Experiment 1A, participants were not required to rate their confidence.

3.2. Results

We conducted a repeated-measures ANOVA with musical training (musician vs non-musician) as a between-participants variable and tonal stability (stable diatonic, unstable diatonic, chromatic tones) and mode (major/minor) as within-participant variables.

The mean brightness ratings are presented in Fig. 6. The main effect of tonal stability was significant, $F_{2,36} = 8.09$, $p = 0.001$, $\eta_p^2 = 0.310$. Pairwise comparisons using Bonferroni corrections revealed that ratings were higher for stable than for unstable diatonic tones, $p = 0.007$, and these were higher than ratings for the chromatic-tone category, $p = 0.038$. The interaction between tonal stability and musical training did not reach significance level, $p = 0.147$, and neither did the three-way interaction, $p = 0.248$.

3.3. Discussion

The results of Experiment 1b replicated the main finding of Experiment 1A: there was a clear correspondence of tonal stability and visual brightness. By

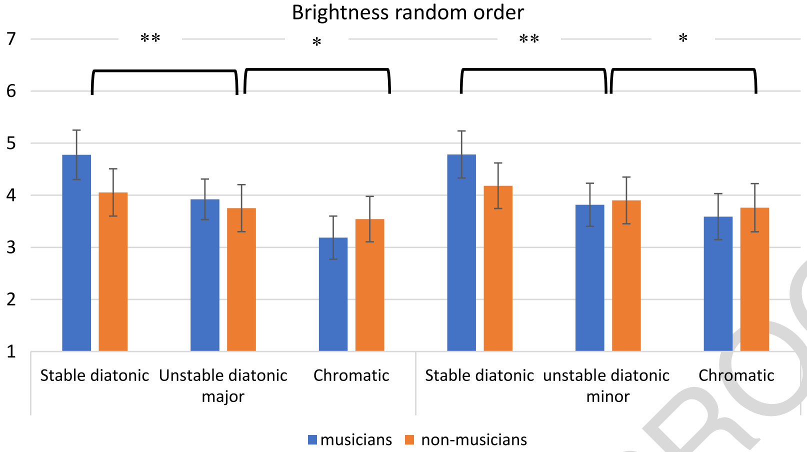

Figure 6. Mean brightness matchings in Experiment 1b (in which brightness levels were presented in randomized order) as a function of tonal stability categories, in major (right panel) and minor (left panel) modes and for musicians (blue bars) and non-musicians (red bars). 1 = darkest circle, 7 = brightest circle.

randomizing the order of brightness levels (rather than presenting them in a monotonically graded order, as in Experiment 1A), we prevented participants from relying on explicit top-down processing. We did not replicate the modulation of this correspondence by musical training, although the numerical trends were in the same direction as in Experiment 1. Two possible interpretations come to mind. One is that randomizing the brightness scale order may have reduced musicians' reliance on conceptual musical knowledge, presumably because in this experiment participants were less able to explicitly associate brightness and tonal stability via a shared scalar representation. The other is that the sample size in the present experiment was smaller than in Experiment 1A ($n = 20$ vs 40, respectively). In any event, the results of this experiment remove any concern that participants' construing of the brightness rating task as an ordinal ranking task accounted for the correspondence of tonal stability and visual brightness observed in Experiments 1A and 1B.

4. Experiment 2

The modified probe–tone paradigm used in Experiment 1 relies on explicit associations between tonal stability and visual brightness. The assignments of different values on a scale (in the GOF task) or along a perceptual dimension (in the brightness-matching task) are based on participants' conscious introspection. To examine whether similar crossmodal correspondences may also be evoked implicitly and unintentionally, a different experimental paradigm is needed.

In Experiment 2, we examined implicit associations between tonal stability and visual brightness by using a variant of the IAT (Greenwald *et al.*, 1998). The IAT was originally designed to assess implicit attitudes — evaluations that occur unintentionally and unconsciously — toward certain population groups. It was later adapted by Parise and Spence (2012) to gage implicit associations between perceptual dimensions in different modalities. Experiment 2 is based on that crossmodal version of the paradigm.

On each trial of our IAT experiment, participants were presented with one of four kinds of stimuli, two auditory and two visual (for examples of stimuli and procedure see Fig. 7 in the Methods section and Supplementary Audio S2 and Supplementary Figs S3 and S4). The auditory stimuli were a tonally stable progression, [A]stable, and a tonally unstable progression [A]unstable (i.e., a key-establishing chord sequence followed by a stable or an unstable tone in that key, respectively). The visual stimuli were a bright and a dark circle, [V]bright, [V]dark, presented on a computer screen. On each trial, participants were asked to categorize a single stimulus (auditory or visual) as quickly and accurately as possible. In each block of trials, two stimuli — one auditory and one visual — were assigned the same response key in a given block of trials. Thus, throughout the block, participants used only two keys, each key paired with one value of the auditory stimuli and one value of the visual stimuli. There were two types of blocks. In congruent blocks, the hypothetically compatible auditory and visual stimuli — [A]stable and [V]bright, [A]instable and [V] dark — were assigned the same response keys, while in incongruent blocks, incompatible stimuli — [A]stable and [V]dark, [A]instable and [V] bright — were assigned to the same response keys. We expected performance in congruent blocks to be better (faster and more accurate) than in incongruent blocks. Specifically, we reasoned that if tonality and brightness are implicitly associated in the same way as they are explicitly associated (Experiments 1A and B), using the same key for [A]stable and [V]bright stimuli and for [A]unstable and [V]dark stimuli should be easier than using different keys.

Crucially, note that this paradigm does not require any explicit judgment of how tightly the auditory and visual stimuli are associated. Such association — whether conscious or unconscious — is orthogonal to the task. Thus, any effects of congruence between tonal stability and brightness on the dependent variables (RT and accuracy) would reflect an unintentional, involuntary association between the two dimensions.

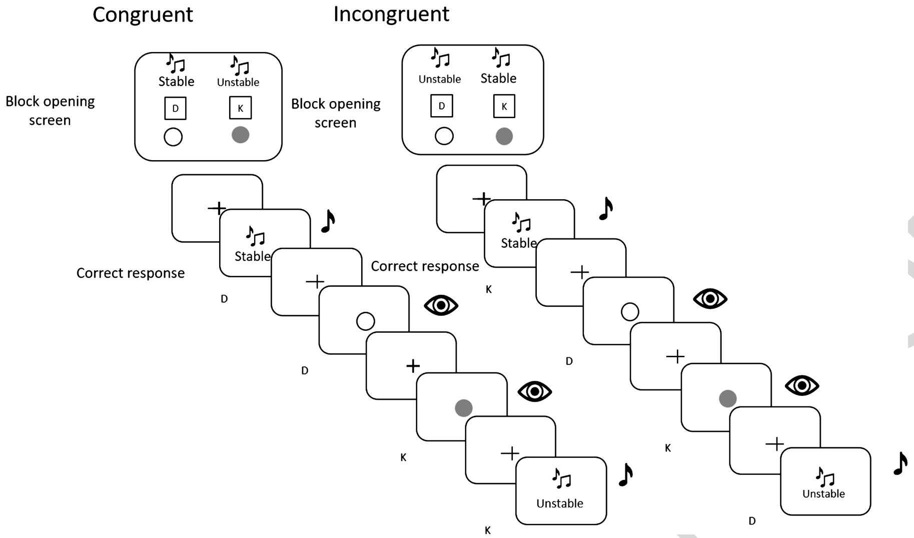

Figure 7. Examples of congruent block (left) and incongruent block (right) trial sequence of Experiment 2. In the beginning of each block, key assignments (K or D) for visual and auditory stimuli were presented to the participants. In this example, for the congruent block (left) both the stable auditory stimulus and the bright visual stimulus were assigned to the D key on the computer keyboard, while the unstable auditory stimulus and the dark visual stimulus were assigned to the K key. For the incongruent block (right), the unstable auditory stimulus and the bright visual stimulus were both assigned the D key, while the stable auditory stimulus and the dark visual stimulus were both assigned the K key. On each trial, participants were presented with a unimodal (visual or auditory) stimulus, and were asked to rapidly press the appropriate key (K or D), as assigned for the specific block.

4.1. Methods

4.1.1. Sample Size Selection

On the basis of the main effect of Parise and Spence (2012) (Exp. 3, pitch height and size), we calculated the sample size required to observe a significant compatibility effect between tonal stability and visual brightness. We conducted this analysis with G*Power (Erdfelder *et al.*, 1996) using an alpha of 0.05, and a power of 0.80. We found the minimum required sample size to be four participants. As we examined possible interactions of this effect with musical training (musicians vs non-musicians) and modality (auditory vs visual), the minimum sample size was 16.

4.1.2. Participants

Thirty undergraduate students from Tel Aviv University served as participants, 15 musicians (8 female, mean age = 22.09, SD = 2.9) and 15 non-musicians (8 female, mean age = 23.48, SD = 1.9). The musicians had an average of 12.29 (SD = 3.4) years of musical experience (with a minimum of eight years) and an average of 6.41 (SD = 2.77) years of music theory studies (with a minimum of five years). They all currently played and performed music. The non-musicians had an average of 0.42 (SD = 1.04) years of musical experience (with a maximum of three years in childhood), and no music theory education; none of them currently played music. All participants were paid 15$ per hour for their participation.

4.1.3. Apparatus and Stimuli

The apparatus was the same as in Experiment 1.

Auditory Stimuli. The auditory stimuli in this experiment were designed to be comparable to the auditory stimuli in Experiment 1A and B (e.g., element and probe, for examples see Supplementary Audio S2 and Supplementary Figs S3 and S4). Since this experiment was speeded, we created a shorter and faster element consisting of a half-cadence (IV–V46–35) — a chord sequence suggesting a subsequent resolution by a stable chord or tone. Each chord of the element sequence lasted for 250 ms separated by 250 ms from the next. Since the IAT paradigm requires a speeded classification between two levels, we used the two probes receiving the highest and lowest brightness ratings in Experiments 1A and 1B. Therefore, this element was followed by a 250 ms tone, either the tonic note (scale degree 1, a tonally stable stimulus) or the raised subdominant (4#, a tonally unstable chromatic note). To minimize possible short-term memory effects, the cadence (played in a three-voice texture) was designed such that neither of the final probe tones (tonic, raised 4th) was included in the chords preceding it. The final tone was octave-doubled, and the pitch direction (up/down) between the last chord in the cadence and the following tone was controlled. Each stimulus was presented in two keys (C major, D-flat major), in a piano timbre (generated by Notion 6 music notation software). The sound examples in Supplementary Audio S2 demonstrate stable and instable auditory stimuli used in this experiment.

Visual Stimuli. The visual stimuli consisted of two circles, one with a brightness level of 24 cd/m^2 (bright) and the other with a brightness level of 1.25 cd/m^2 (dark).

4.1.4. Procedure

For a trial sequence example, see Fig. 7. On each trial, participants were presented with a unimodal stimulus (either auditory or visual). They were asked to classify it using one of two keys (K or D). For the visual stimuli, one key was assigned to the bright circle, and another to the dark circle. For the auditory stimuli, one key was assigned to a stable progression, and the other to an unstable progression. Note that the experimenter did not mention any of these adjectives (dark, bright, stable or unstable) but instead referred to the visual and auditory stimuli as type K and type D according to the participant's key assignment in the practice blocks. The experiment consisted of 24 blocks. All stimuli were presented in all blocks. Each block consisted of 28 consecutive trials. The first four trials served as practice trials and included one stimulus of each type (auditory stable, auditory unstable, visual bright and visual dark). In the following 24 experimental trials, the two visual stimuli were presented six times each, and the four auditory stimuli (stable and unstable in C and C# keys) were presented three times each, all randomly mixed. Each of the four possible stimulus–key pairings (two congruent and two incongruent pairings) was presented in different blocks. There were six blocks for each pairing, resulting in 24 experimental blocks presented in randomly mixed order. Thus, in half of the blocks, the response pairing was congruent (i.e., the same response was associated with the stable auditory stimuli and the bright circle and with the unstable auditory stimuli and the dark circle), whereas in the remaining half of the blocks, the response pairing was incongruent (i.e., the same response was associated with the unstable auditory stimuli and the bright circle and with the stable auditory stimuli and the dark circle).

4.2. Results

The data from one non-musician participant were excluded because his accuracy rate was lower than the group's mean by more than three standard deviations (60% vs $M = 85.33\%$, SD $= 8.54\%$ for non-musicians).

A repeated-measures ANOVA was conducted, with musical training (musician vs non-musician) as a between-participants variable, compatibility (congruent/incongruent) and modality (auditory/visual) as within-participant variables, and mean accuracy rates and reaction times for correct responses as the dependent variables. The mean reaction times and accuracy rates in Experiment 2 are presented in Fig. 8.

4.2.1. Reaction Times

All three main effects were significant. Participants were faster in congruent than in incongruent blocks, $F_{1,27} = 6.511$, $p = 0.017$, $\eta_p^2 = 0.194$, musicians were faster than non-musicians, $F_{1,27} = 6.377$, $p = 0.018$, $\eta_p^2 = 0.191$, and RTs to visual stimuli were faster than RTs to auditory stimuli, $F_{1,27} = 5.076$, $p = 0.033$, $\eta_p^2 = 0.158$. There were no other significant effects, all ps > 0.05. In particular, the compatibility effect did not interact with either musical experience or modality, $F < 1$.

4.2.2. Accuracy Rates

The main effect of compatibility was significant, $F_{1,27} = 28.9$, $p < 0.001$, $\eta_p^2 = 0.517$, with higher accuracy rates in congruent than in incongruent blocks, and so was the main effect of modality, $F_{1,27} = 14.121$, $p = 0.001$, $\eta_p^2 = 0.343$, with higher accuracy for visual than for auditory stimuli. The interaction between modality and musical experience was significant, $F_{1,27} = 10.599$, $p = 0.003$, $\eta_p^2 = 0.282$, indicating that musicians were more accurate

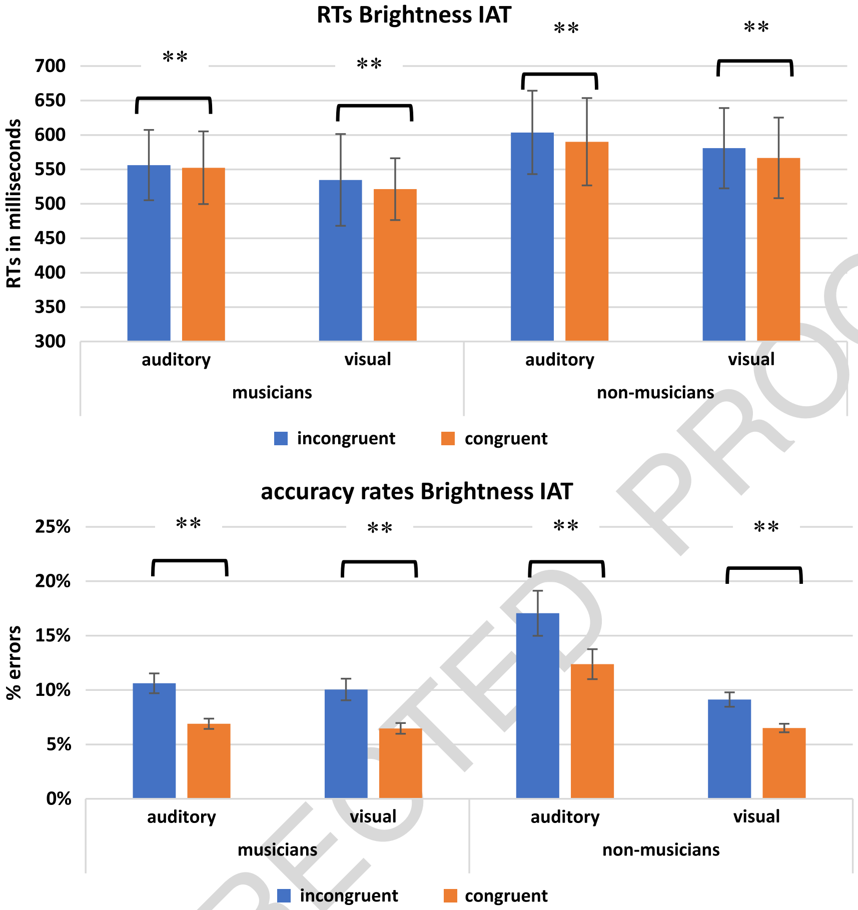

Figure 8. Mean reaction times (upper graph) and accuracy rates (lower graph) in Experiment 2, for congruent (blue) and incongruent (red) blocks, by modality (visual and auditory) for musicians (left) and for non-musicians (right).

at responding to auditory than to visual stimuli. There were no other significant effects, all *ps* > 0.05. Again, the compatibility effect did not interact with either musical experience or modality, both Fs < 1.

4.3. Discussion

The results of Experiment 2 demonstrated a clear correspondence between tonal stability and visual brightness. Performance in congruent blocks, in which tonality-stable auditory stimuli and bright visual stimuli were associated with one key, and unstable auditory stimuli and dark visual stimuli were associated with the alternative key, was significantly better than performance in incongruent blocks. This advantage was observed in both reaction times and accuracy, and for both musicians and non-musicians. The compatibility effect, established with an implicit methodology, confirms the hypothesis that the correspondence between tonal stability and brightness exists at an implicit level and does not necessarily depend on conscious perception. Notably, the compatibility effect was similar in size for musicians and for non-musicians, indicating that at the implicit level, musical training does not strengthen the correspondence.

5. General Discussion

The results of the three experiments reported here suggest that listeners associate tonal hierarchy and visual brightness: tonally stable tones were associated with brighter visual stimuli, and instable tones, with darker stimuli. We explored this CMC using both explicit (modified probe–tone method) and implicit (crossmodal IAT) measures. In Experiments 1A and 1B, participants matched tonally more stable sequences with visually brighter stimuli. In Experiment 2, participants' performance was better (shorter reaction times and higher accuracy) in the congruent condition — when the task required them to use the same motor response to categorize bright visual stimuli and stable auditory stimuli, or dark and unstable stimuli— than in the incongruent condition, in which they used the same motor response to categorize dark visual stimuli and stable auditory stimuli, or bright and unstable stimuli.

Several aspects of our findings suggest that the correspondence between tonal stability and visual brightness does not chiefly rely on intentional application of conceptual musical knowledge. Rather, our results suggest that such correspondences are activated by involuntary, possibly unconscious processes, independent of music-related conceptualization. When listeners were explicitly required to match sequences varying in tonal stability with visual stimuli varying in brightness (Experiments 1A, 1B), the association was observed in both musicians and non-musicians, yet it was stronger for musicians. This finding suggests that conceptual musical knowledge (e.g., identifying a particular stimulus as 'tonally stable' or 'unstable'), which non-musicians lack, strengthens the association between tonal stability and visual brightness, but does not entirely account for the observed correspondence. In direct support of this conclusion, we also found that when the association between tonal stability and visual brightness was gauged using an implicit measure known to reduce the role of consciously mediated decisional processes (IAT, in Experiment 2), musical training did not modulate the strength of the association.

Importantly, while the results of Experiment 1 suggest that CMC between tonal stability and visual brightness do not depend on music-specific conceptualization or on explicit comprehension of tonal hierarchy, these results could still be attributed to a strategic, top-down process. Non-musicians, while unable to categorize auditory stimuli in music-theoretical terms, could still categorize these stimuli along an ad-hoc dimension, possibly based upon implicit grasp of tonal hierarchy (e.g., pleasant/unpleasant, fitting/unfitting) and accordingly, strategically match them with various degrees of brightness (e.g., pleasant → bright).

The results of the IAT task (Experiment 2), however, argue against this possibility: the task required no explicit judgment of audio-visual congruence and any such association was in fact orthogonal to the task. Our results thus suggest that tonality/brightness CMC need not be based on a voluntary or conscious decisional process (see Spence and Deroy [2013)]for similar inferences; Deroy [2019] for relevant discussion of non-conceptual categorization).

Needless to say, these results do not necessarily suggest innate, 'natural' or universal tendencies, since all our participants were exposed to tonal Western music throughout their lives. Rather, they suggest that consistent crossmodal associations between music-syntactic functions and a perceptual dimension (here, visual brightness) may arise implicitly in listeners exposed to music utilizing the relevant syntax (Western tonality), even when these listeners are unaware of these associations and ignorant of the concepts (tonal stability, tonic, etc.) denoting them. Just like listeners' implicit learning of tonal syntax itself, these crossmodal associations may be based on lifetime exposure to tonal music and statistical learning of its connotations. Are universal or innate factors also involved in such associations? Performing similar experiments in a cross-cultural setting may address this intriguing issue.

5.1. Contributions to Crossmodal Correspondence and Music Cognition Research

What may our findings contribute to CMC research, on the one hand, and to the study of music cognition and expression, on the other? For CMC research these findings present a new type of crossmodal correspondence. Rather than examining the associations between basic perceptual dimensions in different sensory modalities (e.g., visual brightness and loudness), we investigated the association between a basic dimension (visual brightness) and a feature generated by a high-level cognitive schema (tonal stability) within a culturally-mediated auditory domain (music). The fact that such correspondences may be featured not only as verbally-mediated, conscious associations (as previous studies, e.g., Arthur [2018] have shown), but also as involuntary implicit associations, independent of conceptual musical knowledge, makes them particularly intriguing as a new avenue for CMC research.

Perhaps most importantly, the present findings open the way for similar research in the domain of language. CMC research in language has so far been conducted mainly in two related contexts: sound symbolism, where the sound of lexical items is associated, via their perceptual connotations, with their meaning (e.g., Walker, 2016b; Walker and Parameswaran, 2019); and as a referential source in speech prosody (e.g., Tzeng *et al*., 2018). This research has typically uncovered associations of basic auditory dimensions, as produced in speech, with basic dimensions in other sensory domains, to suggest connotation or reference (e.g., speakers use higher pitch for novel words denoting brighter colors; Tzeng *et al*., 2018). The type of CMC presented here is substantially different, as is its potential application to language. We suggest that abstract syntactic structures may connote concrete perceptual features across modalities: for instance, a syntactic structure may be associated with brighter or darker luminosity, or with larger or smaller physical size. As is the case with other types of CMC, such associations may not be mere curiosities, but could meaningfully affect information processing, both explicitly and implicitly.

For the study of music cognition, the results presented here suggest that the schemas that govern musical structure (e.g., tonality) may shape the connotative and referential meanings of music by affecting its associations with non-auditory perceptual domains. Indeed, in music theory and related music cognition studies, the notion that structural schemas are implicated in shaping musical meaning is widely accepted (e.g., Huron, 2006; Meyer, 1956). Yet in most such accounts, musical schemas are thought to be involved primarily in introversive musical meanings — 'meanings that are associated with the dynamics of musical materials themselves' (Clarke, 2017), and only indirectly (e.g., as an unexpected event evokes surprise or raises tension) in extroversive meanings 'that point beyond musical materials to a wider world'. Here, we demonstrated a robust association between a central aspect of musical structure (tonal stability) and a basic non-auditory dimension (visual brightness). Moreover, we showed that implicit, non-verbal and unintentional processes underlie — at least in part — the perception of tonally more stable musical events as being 'brighter' than tonally unstable events. While musicians have often alluded to similar associations between tonality and non-auditory perceptual dimensions (Rothfarb, 2001), very few empirical studies have systematically investigated such associations (e.g., Huron, 2006; Arthur, 2018), and none have demonstrated them using implicit, non-verbal experimental measures. The implicit processing of CMC between musical syntax and non-auditory dimensions demonstrated here could also accompany listeners' experience of music in an ecological setting, forming the basis for associations of music with non-auditory domains.

5.2. Some Open Questions

The present study calls for further research to clarify a number of open issues. First, several studies suggest that in musical contexts (and perhaps elsewhere), CMC may be mediated by emotion, such that two dimensions in different perceptual modalities may be associated primarily because both convey a similar emotion, or are associated with the same basic emotional dimension, such as valence (Bhattacharya and Lindsen, 2016; Levitan *et al.*, 2015; Palmer *et al.*, 2013; Whiteford *et al.*, 2018). Indeed, visual lightness and brightness clearly possess emotional connotations, as evidenced not only by conventional verbal metaphors and idioms, but also by a host of experimental research that has established implicit associations between lighter or brighter visual stimuli and positive valence. For instance, positively valenced words were processed faster when printed in white rather than black, while the opposite was true for negative words (Meier *et al.*, 2004). Conversely, positive and negative evaluative words affected brightness perception (Meier *et al.*, 2007). Likewise, tonal stability generates emotional connotations: tonally stable tones or chords are associated with more positive valence (and with specific positive emotions), relative to unstable tones or chords (Arthur, 2018; Huron, 2006; Maimon *et al.*, 2018; Steinbeis *et al.*, 2006). In light of the above, the question arises whether the association between tonal stability and visual brightness is mediated by the emotional connotations of the two dimensions: are tonics brighter because they are more stable and hence 'happier'? Even more intriguingly, one may speculate that emotion may mediate other CMC, involving basic perceptual dimensions not evidently associated with emotion. In other words, it may be the case that all CMCs are mediated by a non-modal process, which, in many cases, may be emotional.

Second, it would be useful to investigate whether the CMC between tonal stability and visual brightness demonstrated here extends to other non-auditory dimensions (e.g., size, spatial height), using the two paradigms proposed here. Such studies (currently in progress in our lab; see Maimon *et al.*, 2017, 2018, 2019 for preliminary reports) should provide a wider picture of the gamut of crossmodal associations of tonal structure, and of the processes underlying them. In particular, they could clarify the role of emotion in the correspondence between tonality and non-auditory dimensions. For example, we showed that while emotional facial expressions (e.g., sad and happy) were associated to music tonality, a dimension that that lacks any distinct emotional connotation (physical size) was not associated to music tonality, either explicitly or implicitly (Maimon *et al.*, 2018).

Finally, while the present study demonstrated implicit associations between tonality and visual brightness, it is not clear whether such crossmodal associations are actually perceptual — whether, for instance, a visual stimulus presented concurrently with a tonally unstable tone would actually be perceived as darker than the same stimulus concurrently presented with a tonally stable tone. A number of empirical studies have demonstrated that some CMC between basic perceptual dimensions do just that (see Spence, 2011; Eitan, 2017, for a survey) — for instance, we perceive 'higher' pitches as emanating from spatially higher locations (Pratt, 1930), and pitch direction affects how we perceive the spatial direction of concurrent visual motion (Maeda *et al.*, 2004). Are tonal relationships capable of generating similar effects — is musical structure capable of actually brightening (and darkening) our world?

Acknowledgements

We thank Roni Granot, Liad Mudrik and Yeshayahu Shen for very useful discussions. We thank Inbal Gur-Arie for her help in performing the experiments.

Funding

This research is supported by an Israel Science Foundation grant (ISF 1920/16) to Zohar Eitan.

Supplementary Material

Supplementary material is available online at:
https://doi.org/10.6084/m9.figshare.12110205

Notes

1. All pitches an octave (or its multiples) from each other belong to the same pitch class (or share the same pitch chroma quality). For instance, the pitches A1 (F0 = 55 Hz), A2 (110 Hz), A3 (220 Hz), A4 (440 Hz), etc., all belong to the pitch class denoted by the note name ‘A-natural.’
2. These correlations included all participants, irrespective of the order in which they took the tasks (GOF first vs brightness first). Correlations were similar in the two groups, although numerical trends were slightly stronger for the participants who performed the brightness task first.

References

Arnheim, R. (1984). Perceptual dynamics in musical expression, *Musical Q.* **70**, 295–309.

Arthur, C. (2018). A perceptual study of scale-degree qualia in context, *Music Percept.* **35**, 295–314.

Bhattacharya, J. and Lindsen, J. P. (2016). Music for a brighter world: brightness judgment bias by musical emotion, *PLoS ONE* **11**, e0148959. DOI:10.1371/journal.pone.0148959.

Bigand, E., Tillmann, B. and Poulin-Charronnat, B. (2006). A module for syntactic processing in music?, *Trends Cogn. Sci.* **10**, 195–196.

Boltz, M. (1989). Perceiving the end: effects of tonal relationships on melodic completion, *J. Exp. Psychol. Hum. Percept. Perform.* **15**, 749–761.

Boulez, P. (1986). *Orientations*. J.-J. Nattiez (Ed.), M. Cooper (trans.). Faber and Faber, London, UK.

Brunel, L., Carvalho, P. F. and Goldstone, R. L. (2015). It does belong together: cross-modal correspondences influence cross-modal integration during perceptual learning, *Front. Psychol.* **6**, 358. DOI:10.3389/fpsyg.2015.00358.

Carlsen, J. C. (1981). Some factors which influence melodic expectancy, *Psychomusicology* **1**, 12–29.

Clarke, E. F. (2017). Music's meanings, in: *The Routledge Companion to Music Cognition*, R. Ashley and R. Timmers (Eds), pp. 523–533. Routledge, New York, NY, USA.

Cuddy, L. L. and Lunney, C. A. (1995). Expectancies generated by melodic intervals: perceptual judgments of melodic continuity, *Atten. Percept. Psychophys.* **57**, 451–462.

Deroy, O. (2019). Categorising without concepts, *Rev. Philos. Psychol.* **10**, 465–478.

Eitan, Z. (2013). How pitch and loudness shape musical space and motion, in: *The Psychology of Music in Multimedia*, S.-L. Tan, A. J. Cohen, S. D. Lipscomb and R. A. Kendall (Eds), pp. 165–191. Oxford University Press, Oxford, UK.

Eitan, Z. (2017). Cross-modal correspondences, in: *The Routledge Companion to Music Cognition*, R. Ashley and R. Timmers (Eds), pp. 213–224. Routledge, New York, NY, USA.

Eitan, Z. and Granot, R. Y. (2006). How music moves: musical parameters and listeners images of motion, *Music Percept.* **23**, 221–248.

Erdfelder, E., Faul, F. and Buchner, A. (1996). GPOWER: a general power analysis program, *Behav. Res. Methods, Instrum. Comput.* **28**, 1–11.

Fétis, F.-J. (1844). *Traité Complet de la Théorie et de la Pratique de l'Harmonie Contenant la Doctrine de la Science et de l'Art*. Maurice Schlesinger, Paris, France.

Globerson, E., Eitan, Z. and Flash, T. (in press). Space, time, and expression in orchestral conducting, in: *Space–Time Geometries in the Brain and Movements in the Arts*, T. Flash and A. Berthoz (Eds). Springer, Berlin, Germany.

Granot, R. and Donchin, E. (2002). Do re mi fa sol la ti — constraints, congruity, and musical training: an event-related brain potentials study of musical expectancies, *Music Percept.* **19**, 487–528.

Greenwald, A. G., McGhee, D. E. and Schwartz, J. L. (1998). Measuring individual differences in implicit cognition: the implicit association test, *J. Pers. Soc. Psychol.* **74**, 1464–1480.

Huron, D. (2006). *Sweet Anticipation: Music and the Psychology of Expectation*. MIT Press, Cambridge, MA, USA.

Koelsch, S. (2011). Toward a neural basis of music perception — a review and updated model, *Front. Psychol.* **2**, 110. DOI:10.3389/fpsyg.2011.00110.

Kohn, D. and Eitan, Z. (2016). Moving music: correspondences of musical parameters and movement dimensions in children's motion and verbal responses, *Music Percept.* **34**, 40–55.

Krumhansl, C. L. (1990). Tonal hierarchies and rare intervals in music cognition, *Music Percept.* **7**, 309–324.

Krumhansl, C. L. (2004). The cognition of tonality — as we know it today, *J. New Music Res.* **33**, 253–268.

Krumhansl, C. L. and Kessler, E. J. (1982). Tracing the dynamic changes in perceived tonal organization in a spatial representation of musical keys, *Psychol. Rev.* **89**, 334–368.

Larson, S. (2004). Musical forces and melodic expectations: comparing computer models and experimental results, *Music Percept.* **21**, 457–498.

Larson, S. and VanHandel, L. (2005). Measuring musical forces, *Music Percept.* **23**, 119–136.

Lerdahl, F. and Jackendoff, R. (1983). *A Generative Theory of Tonal Music*. MIT, Cambridge.

Lerdahl, F. and Krumhansl, C. L. (2007). Modeling tonal tension, *Music Percept.* **24**, 329–366.

Levitan, C. A., Charney, S. A., Schloss, K. B. and Palmer, S. E. (2015). The smell of jazz: crossmodal correspondences between music, odor, and emotion, in: *Cogsci 2015 Proceedings*, pp. 1326–1331. Pasadena, CA, USA.

Maeda, F., Kanai, R. and Shimojo, S. (2004). Changing pitch induced visual motion illusion, *Curr. Biol.* **14**, R990–R991.

Maimon, N., Lamy, D. and Eitan, Z. (2017). Cross-modal correspondences of tonal stability. Talk pesented at the 25th Anniversary Conference of the European Society for the Cognitive Sciences of Music (ESCOM), Ghent, Belgium.

Maimon, N., Lamy, D. and Eitan, Z. (2018). Tonal space, perceived: associations of tonal stability and visual space. Talk presented at the 15th International Conference on Music Perception and Cognition 10th triennial conference of the European Society for the Cognitive Sciences of Music, Graz, Austria.

Maimon, N., Lamy, D. and Eitan, Z. (2019). Associations of tonal stability and visual space: explicit and implicit measures. Talk presented at the 21th conference of the European Society for Cognitive Psychology, Tenerife, SP.

Marks, L. (2004). Cross-modal interactions in speeded classification, in: *The Handbook of Multisensory Processes*, G. Calvert, C. Spence and B. Stein (Eds), pp. 85–106. MIT Press, Cambridge, MA, USA.

Marmel, F. and Tillmann, B. (2009). Tonal priming beyond tonics, *Music Percept.* **26**, 211–221.

Meier, B. P., Robinson, M. D. and Clore, G. L. (2004). Why good guys wear white: automatic inferences about stimulus valence based on brightness, *Psychol. Sci.* **15**, 82–87.

Meier, B. P., Robinson, M. D., Crawford, L. E. and Ahlvers, W. J. (2007). When "light" and "dark" thoughts become light and dark responses: affect biases brightness judgments, *Emotion* **7**, 366–376.

Melara, R. D. and Marks, L. E. (1990). Interaction among auditory dimensions: timbre, pitch, and loudness, *Percept. Psychophys.* **48**, 169–178.

Meyer, L. B. (1956). *Emotion and Meaning in Music*. University of Chicago Press, Chicago, IL, USA.

Naveda, L. and Leman, M. (2010). The spatiotemporal representation of dance and music gestures using topological gesture analysis (TGA), *Music Percept.* **28**, 93–111.

Palmer, S. E., Schloss, K. B., Xu, Z. and Prado-León, L. R. (2013). Music–color associations are mediated by emotion, *Proc. Natl Acad. Sci. USA* **110**, 8836–8841.

Parise, C. V. (2016). Crossmodal correspondences: standing issues and experimental guidelines, *Multisens. Res.* **29**, 7–28.

Parise, C. V. and Spence, C. (2009). 'When birds of a feather fl ock together': synesthetic correspondences modulate audiovisual integration in non-synesthetes, *PLoS ONE* **4**, e5664. DOI:10.1371/journal.pone.0005664.

Parise, C. V. and Spence, C. (2012). Audiovisual crossmodal correspondences and sound symbolism: a study using the implicit association test, *Exp. Brain Res.* **220**, 319–333.

Pratt, C. C. (1930). The spatial character of high and low tones, *J. Exp. Psychol.* **13**, 278–285.

Prendergast, R. M. (1992). *Film Music: a Neglected Art. A Critical Study of Music in Films*, 2nd edn. W. W. Norton & Company, New York, NY, USA.

Rameau, J. P. (1722). *Traité de l'Harmonie Reduite à ses Principes Naturels: Divisé en Quatre Livres*. Jean-Baptiste-Christophe Ballard, Paris, France.

Rothfarb, L. (2001). Energetics, in: *The Cambridge History of Western Music Theory*, T. Christensen (Ed.), pp. 927–955. Cambridge University Press, Cambridge, UK.

Schenker, H. (1906). *Neue musikalische Theorien und Phantasien, Vol. I: Harmonielehre*. J. G. Cotta'sche Buchhandlung Nachfolger, Stuttgart, Germany.

Schoenberg, A. (1978). *Theory of Harmony*. University of California Press, Berkeley, CA, USA.

Shanahan, D. (2017). Musical structure: tonality, melody, harmonicity, and counterpoint, in: *The Routledge Companion to Music Cognition*, R. Ashley and R. Timmers (Eds), pp. 141–151. Routledge, New York, NY, USA.

Shepard, R. N. (1964). Circularity in judgments of relative pitch, *J. Acoust. Soc. Am.* **36**, 2346–2353.

Spence, C. (2011). Crossmodal correspondences: a tutorial review, *Atten. Percept. Psychophys.* **73**, 971–995.

Spence, C. and Deroy, O. (2013). How automatic are crossmodal correspondences?, *Consc. Cogn.* **22**, 245–260.

Spitzer, M. (2003). The metaphor of musical space, *Musicae Sci.* **7**, 101–120.

Steinbeis, N., Koelsch, S. and Sloboda, J. A. (2006). The role of harmonic expectancy violations in musical emotions: evidence from subjective, physiological, and neural responses, *J. Cogn. Neurosci.* **18**, 1380–1393.

Stevens, S. S. (1964). Concerning the psychophysical power law, *Q. J. Exp. Psychol.* **16**, 383–385.

Tzeng, C. Y., Duan, J., Namy, L. L. and Nygaard, L. C. (2018). Prosody in speech as a source of referential information, *Lang. Cogn. Neurosci.* **33**, 512–526.

Walker, P. (2016a). Cross-sensory correspondences: a theoretical framework and their relevance to music, *Psychomusicology* **26**, 103–116.

Walker, P. (2016b). Cross-sensory correspondences and symbolism in spoken and written language, *J. Exp. Psychol. Learn. Mem. Cogn.* **42**, 1339–1361.

Walker, P. and Parameswaran, C. R. (2019). Cross-sensory correspondences in language: vowel sounds can symbolize the felt heaviness of objects, *J. Exp. Psychol. Learn. Mem. Cogn.* **45**, 246–252.

Whiteford, K. L., Schloss, K. B., Helwig, N. E. and Palmer, S. E. (2018). Color, music, and emotion: Bach to the blues, *i-Perception* **9**, 2041669518808535.

Wilson, B., Buelow, G. J. and Hoyt, P. A. (2001). Rhetoric and music, in: *The New Grove Dictionary of Music and Musicians, Vol. 21*, S. Sadie and J. Tyrrell (Eds), pp. 260–275. Oxford University Press, London, UK.