

# Kinetic Cross-Modal Correspondences and Felt (e)Motion in a Novel Set of Musical Stimuli

Anna Kolesnikov<sup>1,\*</sup> , Joshua S. Bamford<sup>2,3,\*</sup> ,  
Eduardo Andrade<sup>4</sup>, Martina Montalti<sup>5,6</sup>, Marta Calbi<sup>6,7</sup>,  
Nunzio Langiulli<sup>6</sup>, Manisha Parmar<sup>8</sup>, Michele Guerra<sup>1</sup>,  
Vittorio Gallese<sup>6,9</sup> and Maria Alessandra Umiltà<sup>8,9</sup>

## Abstract

Embodied music cognition predicts that our understanding of human-made sounds relates to our experience of making the same or similar movements and sounds, which involves imitation of the source of visual and auditory information. This embodiment of sound may lead to numerous kinetic cross-modal correspondences (CMCs). This article investigates music experience in participants with a non-professionally trained music background across three musical dimensions: Contour (Ascending, Descending, Flat), Vertical Density (Low, Medium, High), and Note Pattern (Binary, Ternary, Quaternary). In order that stimuli should reflect contemporary musical usage yet be subject to a high degree of experimental control, 27 ten-second digital piano tracks were created in collaboration with a film composer. In Study 1, participants were asked to rate the stimuli for perceived Direction, Rotation, Movement, and Emotional and Physical Involvement. We test the effects of these factors in terms of the following theories: general and vocal embodied responses to music, the Ecological Theory of Rotating Sounds, and the Shared Affective Motion Experience model of emotion induction. Results for Study 1 were consistent with theories of general and vocal embodied responses to music, as well as with theories of embodied emotional contagion in music. Study 1 also revealed potential confounds in the stimuli, which were further investigated in Study 2 with a new set of participants rating the stimuli for perceived Pitch, Loudness, and Speed. Results for Study 2 served to dissociate intrinsic features of the stimuli from CMCs. Taken together, the two studies reveal a range of embodied CMCs. Although there are limitations to a perceptual study such as this, these stimuli stand to benefit future research in further investigating the embodiment of musical motion.

<sup>1</sup> Department of Humanities, Social Sciences, and Cultural Industries, University of Parma, Parma, Italy

<sup>2</sup> Centre of Excellence in Music, Mind, Body and Brain, Department of Music, Art and Culture Studies, University of Jyväskylä, Jyväskylä, Finland

<sup>3</sup> Social Body Lab, Institute of Human Sciences, School of Anthropology & Museum Ethnography, University of Oxford, Oxford, UK

<sup>4</sup> Royal College of Music, London, UK

<sup>5</sup> Department of Clinical and Experimental Sciences, University of Brescia, Brescia, Italy

<sup>6</sup> Department of Medicine and Surgery, Unit of Neuroscience, University of Parma, Parma, Italy

<sup>7</sup> Department of Philosophy "Piero Martinetti," University of Milan, Milan, Italy

<sup>8</sup> Department of Food and Drug, University of Parma, Parma, Italy

<sup>9</sup> Department of Art History and Archaeology, Columbia University, New York, USA

\*These authors contributed equally.

## Corresponding authors:

Anna Kolesnikov, Department of Humanities, Social Sciences, and Cultural Industries, University of Parma, Parma, Italy.

Email: [anna.kolesnikov@unipr.it](mailto:anna.kolesnikov@unipr.it)

Joshua S. Bamford, Centre of Excellence in Music, Mind, Body and Brain, Department of Music, Art and Culture Studies, University of Jyväskylä, Jyväskylä, Finland; Social Body Lab, Institute of Human Sciences, School of Anthropology & Museum Ethnography, University of Oxford, Oxford, UK.

Email: [joshua.s.bamford@jyu.fi](mailto:joshua.s.bamford@jyu.fi)



Creative Commons CC BY: This article is distributed under the terms of the Creative Commons Attribution 4.0 License (<https://creativecommons.org/licenses/by/4.0/>) which permits any use, reproduction and distribution of the work without further permission provided the original work is attributed as specified on the SAGE and Open Access page

(<https://us.sagepub.com/en-us/nam/open-access-at-sage>).

## Keywords

Contour, cross-modal correspondences, Ecological Theory of Rotating Sounds, embodied music cognition, loudness, mimetic subvocalization, note pattern, pitch height, SAME model, speed

Submission date: 27 August 2021; Acceptance date: 11 October 2023

## Introduction

Music does not re-present anything ... the feeling is presented – enacted – in the felt experience of the listener. To hear the music is just to be moved and to feel in the precise way that is defined by the patterns of the musical motion. (Mark Johnson, 2007, p. 461)

Since the time of Aristotle in his *Problemata*, the perceived “motion” of music has been of significant interest to philosophers and musicologists (see Hanslick, 1891; Pratt, 1931/1958; Schenker, 1906/1954; Truslit, 1938/Repp, 1992; Zuckerkandl, 1956). There is much debate as to whether the motion perceived in music is real (Todd, 1999), virtual (Clarke, 2001), or merely a persistent illusion (Gjerdingen, 1994), but as Godøy and Leman (2010) note, music is fundamentally “a combination of sound and movement,” and the meaning of music for humans derives from this very combination. Progress in the behavioral and neural sciences has provided empirical evidence for the role of corporeal engagement in shaping musical experience (Leman, 2007), as well as for the link between embodied perception and felt emotion (Molnar-Szakacs & Overy, 2006; Overy & Molnar-Szakacs, 2009). Although embodied musical metaphors have been discussed previously (see Lakoff & Johnson, 1998/2008; Scruton, 1997), neurological mechanisms behind these have only been suggested more recently (see below). Such studies provide key support to the framework of embodied music cognition (Cox, 2001, 2011, 2017; Godøy, 2003; Godøy & Leman, 2010; Leman, 2007), which posits that musical meaning may begin with the imitation of musical sounds and of the physical exertions that they produce using the body.

The present article examines how formal acoustic features impact perceived motion, as experienced by listeners without specialist musical training, through simulatory cross-modal correspondences (CMCs), felt effects, and felt emotion. Specifically, using behavioral measures, we investigate the phenomenon of kinetic CMCs in relation to Pitch Contour, Vertical (or Harmonic) Density, and Note Pattern, as well as the connection between kinetic CMCs and felt emotion. In doing so, we draw upon multiple theoretical frameworks that will be discussed further.

## CMCs in Music

A perceptual phenomenon observed across many different domains, CMCs refer to “systematic associations found across seemingly unrelated features from different sensory

modalities” (Parise et al., 2016, in Eitan, 2017, p. 213). Certain CMCs are linked to perceived motion (i.e., kinetic CMCs) and thus are a form of embodied music reception. Most studies have asked participants for subjective ratings; however, there are also examples where participants have had motor responses recorded during music listening, finding a strong correspondence between pitch and vertical motion (Godøy et al., 2006; Kelkar & Jensenius, 2019; Krantz et al., 2006; Küssner & Leech-Wilkinson, 2014; Küssner et al., 2014; Nymoen et al., 2011). The origins of this perceptual experience may arise from CMCs formed through experience with ecological stimuli (e.g., Hansen & Huron, 2019), or from internal stimuli, using the self (body/voice) as a reference point (e.g., Cox, 2017). In either case, this may stem from learnt associations with the functional movements that may be required to produce the perceived sound.

The extent to which CMCs are learnt or innate is a matter of debate. Mondloch and Maurer (2004) suggest that some CMCs are likely innate due to remnant cross-modal neural connections: specifically for pitch, size, and brightness. On the other hand, Spence and Deroy (2012) argue that most CMCs are likely learnt through associations between environmental stimuli. Developmental studies suggest that at least some cross-modal associations are present from a very early age (Kohn & Eitan, 2009; Walker et al., 2010), although some associations may be weaker in children than in adults (Eitan & Tubul, 2010), suggesting an effect of learning.

Furthermore, the kinetic qualities of music are made evident in embodied metaphors that may be acquired through linguistic associations. Examples from Italian, commonly used in Western music notation, include: *crescendo* (crescere = to rise), *decrescendo* (decrescere = to fall), *staccato* (staccare = to detach), *legato* (legare = to link), *forte* (strongly), and *piano* (softly). These modes of movement are so deeply encoded in our language that “it is almost impossible not to think of music in those terms” (Stern, 2010, p. 57). Pitch contours are often described vertically in many languages, with “high” pitch usually referring to a greater frequency, although this is not universal (Žuvela & Anić, 2019), but embodied linguistic metaphors may vary greatly depending on language and culture (e.g., Dolscheid et al., 2013). Similarly, experience with an instrument through musical training may shape some spatial CMCs; for instance, pianists are more likely to map pitch horizontally, according to the keyboard (Stewart et al., 2004). Therefore, many CMCs are likely acquired through enculturation, or through exposure to ecological stimuli, rather than being innate.

## CMCs and Rotating Sounds

Applying the Gibsonian perspective (i.e., that an analysis of the environment is crucial to explaining behavior) (Gibson, 1966, 1979) to the study of music, ecological acoustics studies the relationship between human beings and their environment as mediated through sound. It posits that listeners acquire practical knowledge from sounds in their acoustic surroundings rather than abstract concepts such as frequency, duration, and intensity. Gjerdingen (1994) provides an account of “apparent,” or illusory, motion in music through an analogy with apparent motion in vision. However, in the apparent movement between two tonal events in a melody, it is less clear “what is moving and where that motion takes place” (Gjerdingen, 1994, p. 336). Hansen and Huron (2019) suggest that one such phenomenon is the perception of triplet rhythms. Although music scholars have long been aware of the sensation of rotation evoked in listeners, the authors observe that no theory has been proposed to account for this association. Furthermore, very few theories of CMCs with metrical hierarchy have been proposed (Eitan, 2017). Drawing from the domain of ecological acoustics, the authors propose an Ecological Theory of Rotating Sounds (EToRS), which maps fluctuations in loudness (i.e., accents) to trajectories of rotating sound sources. According to the theory, in the absence of an existing metrical context, the listener tends to hear the loudest event as marking the downbeat (Hansen & Huron, 2019). Thus, if the listener is situated outside the trajectory of the rotating sound, Duple Note Patterns (two beats per measure) reduce to pendular motion, whereas Triple Note Patterns (three beats per measure) reduce to spinning movement, and for Quadruple (four beats per measure) and Quintuple Note Patterns (five beats per measure) it is ambiguous as to whether the motion is reduced to spinning or pendular motion.

## Pitch CMCs and the Human Voice

While some CMCs are developed through observing common environmental occurrences, others may arise from experience with internal processes. According to Cox (2017), the voice plays a key role in embodied responses to kinetic CMCs for pitch, which the author denotes through three notions: 1) mimetic subvocalization, 2) “greater is higher,” and 3) “voice as source domain”. Mimetic subvocalization refers to the covert imitation of sounds (such as another’s spoken words or singing) using specifically the voice (p. 29). The “greater is higher” metaphor refers to the increased effort, by way of mimetic subvocalization, in mimicking higher pitch (p. 88). Finally, “voice as source domain” refers to the notion that our voice, our “first instrument,” is a gateway to embodied musical meaning as it makes sense of pitch height in other instrumental sounds (p. 96).

For most people, the voice is the first and most common means of auditory communication and of understanding the

vocal expressions of others (Bannan, 2019, p. 4; Tyack, 2016). Mimetic vocalization and subvocalization (in imitation of voices and other sounds) are also associated with varying muscle tension and effort (Brodsky et al., 2008), which may be the origin of the association of pitch height to motion in vertical space. As Sessions observes (1941, p. 108), from childhood we learn that if we raise our voice, we often end up raising pitch height—they go hand in hand. Studies with adults also find that tension around the neck and shoulders increases with rising pitch (Pettersen et al., 2005). The “greater is higher” metaphor may thus be a kind of CMC for pitch height and intensity that in turn may influence felt emotion. As the goal-oriented nature of music is often structured on climactic moments composed of melodic high-points (Eitan, 1997), and these high-points often correlate with greater pitch height and greater acoustic strength (i.e., volume), the sense of “achievement” and emotional climax in music is amplified. For example, when listening to “Nessun dorma” from Puccini’s *Turandot*, one may tacitly mimic what it would be like to perform the “physical-artistic accomplishment” of sustaining the high B of the final “Vincerò!” (Cox, 2017, p. 91), and by way of this mimicking, be “moved” emotionally. Indeed, all singing involves activation of the abdominal muscles, shaping musical experience by way of correlations between abdominal exertions (tightness and relaxation) and emotional states (Del Negro et al., 2018). This learnt association between pitch height and effort may contribute to the CMC between pitch and vertical motion that has been observed in previous literature. Furthermore, this increase in perceived and felt effort could contribute to an increased intensity of felt emotion in response to a musical stimulus.

## Motor Mirroring and Emotion

Stemming from the discovery of mirror neurons in macaques (Di Pellegrino et al., 1992; Gallese et al., 1996; Rizzolatti et al., 1996) and the theory of embodied simulation theory that followed (Gallese, 2005, 2014; Gallese & Sinigaglia, 2011), a growing body of neuroimaging evidence substantiates the claim that action execution and perception are linked in humans (see Gallese, 2009). As neuroimaging techniques have made it possible to investigate the role of sensorimotor engagement in musical experience, multiple studies demonstrate that motor brain areas are activated during passive music listening (Callan et al., 2006; Chen et al., 2008a, 2008b; Gordon et al., 2018; Halpern & Zatorre, 1999; Halpern et al., 2004; Hickok et al., 2003; Zatorre et al., 1996) as well as during beat entrainment (Grahn & Brett, 2007; Nistri et al., 2006). Such motor responses to music perception may be integral to the perception and elicitation of emotion in music (Hodges, 2009; Leman, 2007). This rests on the theory that empathy and the perception of emotion are embodied processes (Iacoboni, 2009). Indeed, studies have found that music-induced movement may be associated with

heightened positive affect during music listening (Saarikallio et al., 2013).

The Shared Affective Motion Experience (SAME) theory, an extension of the embodied music cognition framework, suggests that perception of emotion is fundamentally based on the perception of movement, and that the implied movement in music leads to the perception of emotion (Overy & Molnar-Szakacs, 2009). The SAME model may underpin the emotional contagion process in music-evoked emotion, which Juslin and Västjäll (2008) suggest may involve motor processing. Subsequent research has found that both music and movement that convey similar emotional content also share common features (Sievers et al., 2013). This also relates to Laban's theory of shape/effort and movement analysis (Laban, 1947), which links dynamic movement with expressive qualities such as flow, space, time, shape, and, notably, weight and tension, and features prominently in research on music-induced body movements (Maes et al., 2014). Indeed, Laban analysis has been used to identify basic emotions in movement (Shafir et al., 2016). Furthermore, there is some evidence to suggest that people with higher Trait Empathy may have a greater embodied response to music (Bamford & Davidson, 2019; Moorthigari et al., 2021; Wallmark et al., 2018), further supporting a relationship between empathy and motor mirroring. It is important to acknowledge that perceived and felt emotions may arise through distinct systems (Gabrielsson, 2002; Timmers, 2017), as people can usually identify emotional content within the music, without necessarily feeling the identified emotion (Juslin & Laukka, 2004). The emotions identified may be treated either as discrete categories, or as two-dimensional emotional space consisting of valence and arousal (Eerola & Vuoskoski, 2011). The present study is only concerned with the strength of embodied perception involved in emotional contagion, rather than any specific emotional responses, per se. As such, the Emotional Involvement measure in this study asked particularly about the intensity of felt emotion.

### Research Aims

As described previously, there are numerous studies on the cross-modal interactions of music in terms of association with physical (i.e., external) space and bodily motion. Furthermore, there is considerable evidence, and a strong conceptual framework, for the claim that we understand human movement and human-made sounds in relation to our own bodily experience, grounded in our own body's ability to imitate the source of visual and auditory information to make similar sounds. Thus, the aims of the current article are to investigate the impact of multiple formal musical features on measures of perceived musical motion and involvement in a population of individuals with no professional or semi-professional musical training. Non-professional musicians were tested because extensive

musical training may result in learnt motor-CMCs associated with an instrument (Stewart et al., 2004).

A novel, naturalistic set of musical stimuli was developed in collaboration with a film composer, targeting perception of motion. The stimuli developed here were systematically manipulated across the following factors: Contour, Vertical Density, and Note Pattern. Contour refers to the direction of melodic movement over the course of the stimulus and has three levels: Ascending, Flat, and Descending. In the context of this study, Vertical Density refers to harmonic density, through the addition of pedal notes and melodic lines. Finally, three different metrical hierarchical structures were included as part of the Note Pattern variable: Duple, Triple, and Quadruple. Other features such as dynamics (*mf*) and tempo (120 bpm) were kept constant. The sound source was also held stationary, and other acoustic features were carefully controlled. In addition, the expressiveness of performance was held constant between the stimuli, so as to only measure the effect of musical structure (Timmers, 2017). It is worth noting that most previous CMC research has used single-tone stimuli or single melodic lines (Eitan et al., 2014), while our stimuli were naturalistic and dynamic. There are limited examples of musical stimulus sets that aim to find a balance between experimental control and maintaining naturalistic aesthetic qualities (e.g., Clemente et al., 2020). Crucially, there must be a theoretical rationale behind any stimulus set, and the musical features that were manipulated within these stimuli were specifically chosen to test the target theories discussed earlier.

These stimuli were used across two behavioral studies that will be described in the following text. Study 1 aimed to test the primary hypotheses. This revealed some potential confounds in the stimuli, which were further investigated in Study 2 by asking a new sample of participants about the musical features of the stimuli. The results of both studies will then be discussed, highlighting the need for future research and ways in which these stimuli could be adapted or improved for investigating the target theories, while also highlighting the need for research that involves controlled musical stimuli such as these.

### STUDY I

The three factors in the stimuli (Contour, Vertical Density, and Note Pattern) were investigated with respect to subjective ratings. Participants were asked to make an evaluation in terms of perceived motion of the music (Direction, Rotation, and Movement), and in terms of the experiences the music evoked in themselves (Physical and Emotional Involvement). By "perceived musical motion," including Direction and Rotation as more specific kinds of Movement, we refer to what Clarke (2001) denotes as "virtual" motion during musical experience, based upon the notion that music evokes a "virtual person" in the mind of the listener (Watt & Ash, 1998). Of these, Rotation may be the most illusory while perceived

Direction may be related to internal processes (i.e., embodied vocal responses). By “involvement” we refer to participants’ felt responses to the music, including both Physical and Emotional Involvement. The study was conducted only with Italian-speaking participants, and so cannot investigate any possible effects of language, where previous studies have found that existing linguistic associations may influence CMC in music perception (Dolscheid et al., 2013). Specifically, we hypothesize that:

1. Participants will correctly identify the Direction of the musical stimuli (i.e., Ascending, Descending, or Flat), based on observed CMCs between pitch contour and vertical movement previously noted in the literature (Küssner & Leech-Wilkinson, 2014);
2. Triple Note Pattern will be perceived as evoking more Rotation than Quadruple and Duple conditions, in accordance with the EToRS (Hansen & Huron, 2019);
3. Ascending Contour will be perceived as evoking more Movement, Physical Involvement, and Emotional Involvement than Descending and Flat conditions, according to Cox’s (2017) theories on embodied vocal responses and the principles of the SAME theory (Overy & Molnar-Szakacs, 2009);
4. High Vertical Density will evoke more perceived Movement, Physical Involvement, and Emotional Involvement compared to Medium and Low conditions, due to the increased density of musical events, according to general theories of embodied responses to music (Godøy, 2003; Leman, 2007) and SAME theory; and
5. Emotional Involvement will correlate with perceived Movement and Physical Involvement, according to the principles of the SAME theory. Although we do not use physiological measures in this article, we test the theories of general and vocal embodied responses indirectly using behavioral measures.

We also expected that there may be some individual differences on the basis of Empathy, Affect, and Motor Imagination, so a battery of questionnaires was included to investigate these factors. Empathy was assessed in all participants using the Interpersonal Reactivity Index (IRI; Albiro et al., 2006; Davis, 1980). Motor imagination was assessed in all participants using the Vividness of Movement Imagery Questionnaire-2 (VMIQ-2; Roberts et al., 2008). Affective and Reactive attitudes to music were assessed in all participants using the Brief Musical Experience Questionnaire (BMEQ; Werner et al., 2006) (see Supplementary Materials 1 for details).

## Materials and Methods

### Participants

Participants were recruited through opportunity sampling using Facebook, which filtered individuals for age, residence in Parma, and a non-professionally trained musical background. Participants were further screened for musical training prior to participation based on a screening questionnaire, with all individuals having no more than 6 years of music classes within the school curriculum or private musical tuition (including voice), with the exception of two participants, with an average total number of years of musical training of  $2.5 \pm 3.73$  standard deviation (SD) years. None of the participants are professional or semi-professional musicians. In total, 30 healthy volunteers of Italian nationality took part in the study: 16 females and 14 males, mean age 28.76 ( $SD=4.64$ ,  $min=18$ ,  $max=35$ ). All participants reported having normal hearing and normal or corrected-to-normal visual acuity. All participants were either right-handed or ambidextrous (as determined by the Edinburgh Handedness Test; Oldfield, 1971). Power was calculated *a posteriori* by means of G\*Power 3.1 (Faul et al., 2007) using the linear multiple regression: a random model to test for a linear mixed effect model for each dependent variable. With a Cohen’s F effect size equal to 0.4 (medium effect size), an alpha level of .05, three predictors, and a total sample size of 30 resulted in an actual power of  $>0.9$ . All participants provided written informed consent to participate in the study, which was conducted in accordance with the Declaration of Helsinki (Ndebele, 2013), complying with the ethical standards of the Italian Board of Psychologists as well as the Ethical Code for Psychological Research of the Italian Psychological Society and the Ethical Committee of the Area Vasta Emilia Nord (AVEN).

### Stimuli Creation

In collaboration with a film composer, musical variations of melodic motion were composed. A selection of 10-s digital piano tracks were created with a digital audio workstation (DAW) using Cubase Pro software, with the piano sample from Garritan/Abbey Road CFX Grand. The tracks were created in the following modalities: melodies with Ascending, Descending, or Flat Contour, with Duple, Triple, or Quadruple Note Pattern, and with Low, Medium, or High Vertical Density (Table 1), thus following a  $3 \times 3 \times 3$  factorial design for a total of 27 tracks. The tracks were composed on a digital piano due to the greater ease of gradually integrating Vertical Density through use of the left and right hands. All notes were played with a certain degree of human variability in their dynamics in order to increase likeness with the dynamics of how human music is usually performed (tempo = 120, all excerpts and dynamics were played at *mf* as an average; see Supplementary Materials (2) for stimuli partitions). In order to prioritize the investigation

of perceived motion, the composer avoided the explicit expression of “valence” through key in the compositions, striving for “neutrality.” All tracks were balanced for volume and melodic structure and formatted as waveform audio files (WAV) with a sampling rate of 44.1 kHz and 16 bits per sample. (For analysis of the stimuli, see subsections “Pitch” and “Loudness” in Study 2 “Results” section. Full stimulus set is available and open access.)

### Procedure

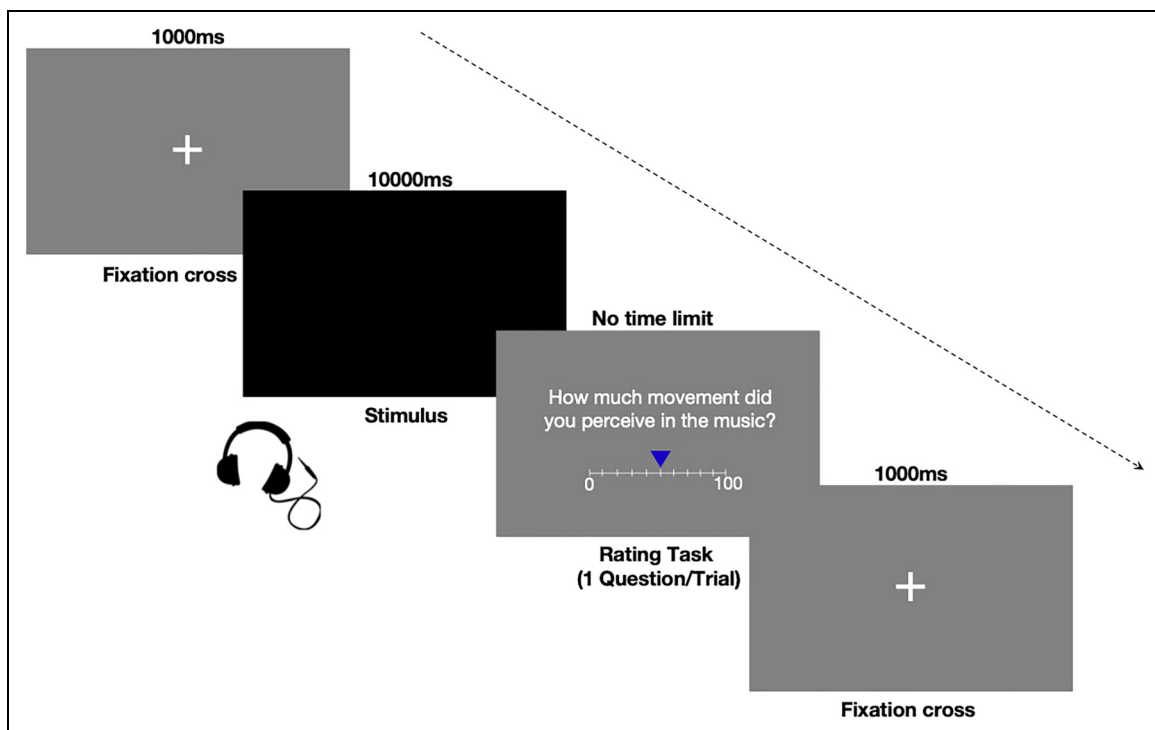
Upon arrival, participants were asked to make themselves comfortable and were given instructions about the study.

**Table 1.** Stimuli/conditions.

Stimulus	Contour	Note Pattern	Vertical Density
1–9	Ascending	Duple	L, M, H
		Triple	L, M, H
		Quadruple	L, M, H
10–18	Descending	Duple	L, M, H
		Triple	L, M, H
		Quadruple	L, M, H
19–27	Flat	Duple	L, M, H
		Triple	L, M, H
		Quadruple	L, M, H

Note: Total number of stimuli: 27. L: Low, M: Medium, H: High. For examples, see: “Ascending, Quadruple, High”; “Descending, Triple, Middle”; Flat, Duple, Low”. Musical scores may be found in Supplementary Materials 2.

The experimental session consisted of two different and randomized phases. In the first phase, participants were asked to fill out a series of questionnaires (for more information, see the Supplementary Materials 1). In the second phase, participants were asked to perform a computer task in which the 27 audio tracks were presented in randomized order. In each trial, a fixation cross was presented for 1,000 ms, the audio stimulus was presented for 10,000 ms (with a black screen), and afterward a question was presented with no time limit (Figure 1). Each stimulus was followed by one of the following questions, in randomized order: 1) “In what direction was the music moving?” 2) “How much rotation did you perceive in the music?”; 3) “How much movement did you perceive in the music?” 4) “How physically involved did you feel?” and 5) “How emotionally involved did you feel?” Participants were asked to listen to the stimuli and answer the questions as quickly and as accurately as possible (no specific time limit was given), using the mouse to move a blue cursor (positioned at 50) along a visual analog scale (VAS) ranging from 0 (very little) to 100 (very much). Only for the “Direction” question, responses tending toward 0 signified “downward” and responses tending toward 100 signified “upward,” with 50 signifying “no change in direction.” The study design is the following: 3 Contours (Ascending, Descending, Flat) \* 3 Note Patterns (Duple, Triple, Quadruple) \* 3 Vertical Densities (High, Medium, Low), for a total of 27 conditions. Each experimental condition was repeated 10 times (1 question/stimulus presentation × 5 questions × 2 repetitions), for a total of 270 presented trials. Before carrying



**Figure 1.** Example of Study I experimental trial. Components: fixation cross frame (1,000 ms), stimulus frame (10,000 ms) and rating task (no time limit). Experiment was created using Psychopy 3.0.

out the experimental procedure participants were presented with a brief training phase to become accustomed with the task. After the experimental session participants were asked to fill out a short debriefing survey about their experience. The experimental session was conducted in a quiet room, on a screen positioned approximately 60 cm from the participant and using Sony WH1,000XM2 noise cancelling headphones at a volume that was comfortable for the participants. The experimental task was programmed using Psychopy 3.0 software (Peirce et al., 2019).

## Analysis

In order to investigate whether VAS ratings were modulated by Contour, Vertical Density, and Note Pattern, a linear mixed effect analysis was carried out for each dependent variable. Participants' ratings were entered as dependent variables (Direction, Rotation, Movement, Emotional Involvement, and Physical Involvement). Contour (three levels: Ascending, Descending, Flat), Vertical Density (three levels: Low, Medium, High), and Note Pattern (three levels: Duple, Triple, Quadruple) were entered as independent fixed variables, and participant intercepts were entered as random effects. Tukey's test was used for post hoc comparisons among means, where a Type 1 error probability of less than 5% ( $p < .05$ ) was considered significant in a population sample size of  $N = 30$ . Ratings for Movement, Rotation, Physical Involvement, and Emotional Involvement were averaged across conditions for each participant for the purpose of calculating Spearman correlations between participants' ratings of perceived Movement and embodied experience during passive music listening and participants' questionnaire scores. Averaged ratings for perceived Direction, Rotation, Movement, and Physical Involvement were correlated with participants' scores for the Internal Visual Imagery,

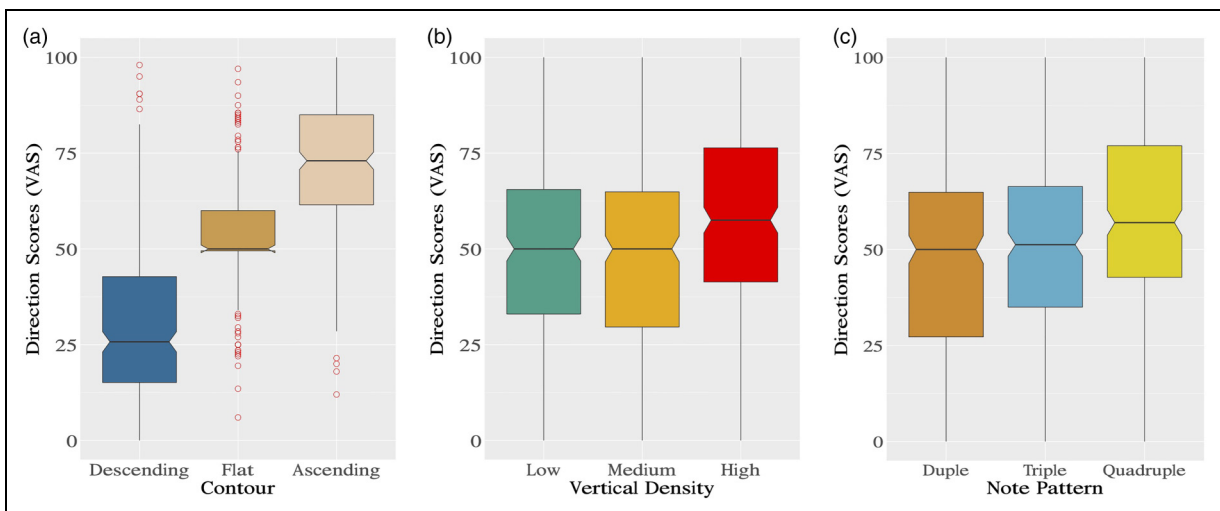
External Visual Imagery, and Kinesthetic Imagery subscales of the VMIQ-2 and the Reactive subscale of the BMEQ (i.e., motile responses to music), and averaged ratings for Emotional Involvement were correlated with participants' scores for the Empathic Concern subscale of the IRI (i.e., empathic personality traits) and the Affective subscale of the BMEQ (i.e., affective responses to music). In addition, ratings for Direction, Movement, and Physical Involvement were correlated with ratings for Rotation, and ratings for Emotional Involvement were correlated with ratings for Direction, Rotation, Movement, and Physical Involvement. The critical probability values for multiple comparisons were corrected using the Bonferroni method. All analyses were performed using R software (R Core Team, 2019), lme4 (Bates et al., 2015), ordinal (Christensen, 2019), effects (Fox, 2003), and emmeans (Lenth et al., 2020) packages; for data visualization, the ggplot2 package was used (Wickham, 2016).

## Results

Estimated marginal means calculated from the linear mixed effects model for all conditions, and test statistics for all Tukey's post hoc comparisons for interaction effects, may be found in Supplementary Materials 3 and 4, respectively.

### Direction

The model explained 55.76% of the variance in Direction ratings, taking into account the random effects ( $R_m^2 = 0.53$ ,  $R_c^2 = 0.56$ ). The model revealed a significant main effect of Contour ( $\chi^2_{(2)} = 869.62$ ,  $p < .0001$ ), showing that participants perceived Ascending to be moving in a higher Direction with respect to Descending and Flat, and Descending in a lower Direction than Flat (Figure 2[a]). A significant main effect for Vertical Density was found



**Figure 2.** Direction. Boxplots depicting mean VAS ratings for Direction with respect to the main effects: (a) Contour, (b) Vertical Density, and (c) Note Pattern ( $N = 30$ ) (see text for significant results).

**Table 2.** Tukey's post hoc comparisons – effect of Contour, Vertical Density, and Note Pattern on perceived direction.

		Mean Difference	SE	<i>t</i> (754)	<i>p</i>
<i>Contour</i>					
Ascending	Descending	42.1	1.43	29.45	< .0001 ***
	Flat	19.1	1.43	13.33	< .0001 ***
Descending	Flat	-23.0	1.43	-16.11	< .0001 ***
<i>Vertical Density</i>					
High	Medium	8.56	1.43	5.99	< .0001 ***
	Low	7.02	1.43	4.91	< .0001 ***
Medium	Low	-1.54	1.43	1.08	.53
<i>Note Pattern</i>					
Quadruple	Triple	4.73	1.43	3.31	.003 **
	Duple	9.02	1.43	6.31	< .0001 ***
Triple	Duple	4.29	1.43	3.00	.008 **

$p < .05 = *$ ,  $p < .01 = **$ ,  $p < .001 = ***$

( $\chi^2_{(2)} = 40.76$ ,  $p < .0001$ ), showing that participants perceived High Vertical Density melodies to be moving in a higher Direction than Medium and Low (Figure 2[b]). A significant main effect for Note Pattern was found ( $\chi^2_{(2)} = 39.83$ ,  $p < .0001$ ), showing that participants perceived Quadruple to be moving in a higher Direction than Triple and Duple, and Duple moving in a lower Direction than Triple (Figure 2[c]; for Tukey's post hoc comparisons for main effects, see Table 2).

### Rotation

The model explained 44.83% of the variance in Rotation ratings, taking into account the random effects ( $R_m^2 = 0.29$ ,  $R_c^2 = 0.45$ ). The model revealed a significant main effect of Contour ( $\chi^2_{(2)} = 293.57$ ,  $p < .0001$ ), showing that participants perceived more Rotation in Ascending than in both Descending and Flat, and in Descending than Flat (Figure 3[a]). A significant main effect for Vertical Density was found ( $\chi^2_{(2)} = 33.83$ ,  $p < .0001$ ), showing that participants perceived more Rotation in High Vertical Density than Medium and Low (Figure 3[b]). A significant main effect for Note Pattern was found ( $\chi^2_{(2)} = 54.09$ ,  $p < .0001$ ), showing that participants perceived more Rotation in Quadruple than in Triple and Duple, and in Duple less than in Triple (Figure 3[c]). Significant interaction effects for Vertical Density \* Contour ( $\chi^2_{(4)} = 21.15$ ,  $p < .001$ ) and Note Pattern \* Contour ( $\chi^2_{(4)} = 12.56$ ,  $p = .014$ ) were found, (Figures 3[d] and 3[e], respectively; for Tukey's post hoc comparisons for main effects, see Table 3).

### Movement

The model explained 40.92% of the variance in Movement ratings, taking into account the random effects ( $R_m^2 = 0.27$ ,  $R_c^2 = 0.41$ ). The model revealed a significant main effect of Contour ( $\chi^2_{(2)} = 141.89$ ,  $p < .0001$ ), showing that participants perceived more Movement in Ascending than in Descending and Flat, and Descending

more than Flat (Figure 4[a]). A significant main effect for Vertical Density was found ( $\chi^2_{(2)} = 55.16$ ,  $p < .0001$ ), showing that participants perceived more Movement in High than Medium and Low (Figure 4[b]). A significant main effect for Note Pattern was found ( $\chi^2_{(2)} = 136.18$ ,  $p < .0001$ ), showing that participants perceived more Movement in Quadruple than in Triple and Duple, and in Duple less than in Triple (Figure 4[c]). Significant interaction effects were found for Note Pattern \* Vertical Density ( $\chi^2_{(4)} = 9.75$ ,  $p = .045$ ) and for Contour \* Vertical Density ( $\chi^2_{(4)} = 15.37$ ,  $p = .004$ ), (Figures 4[d] and 4[e], respectively; for Tukey's post hoc comparisons for main effects, see Table 4).

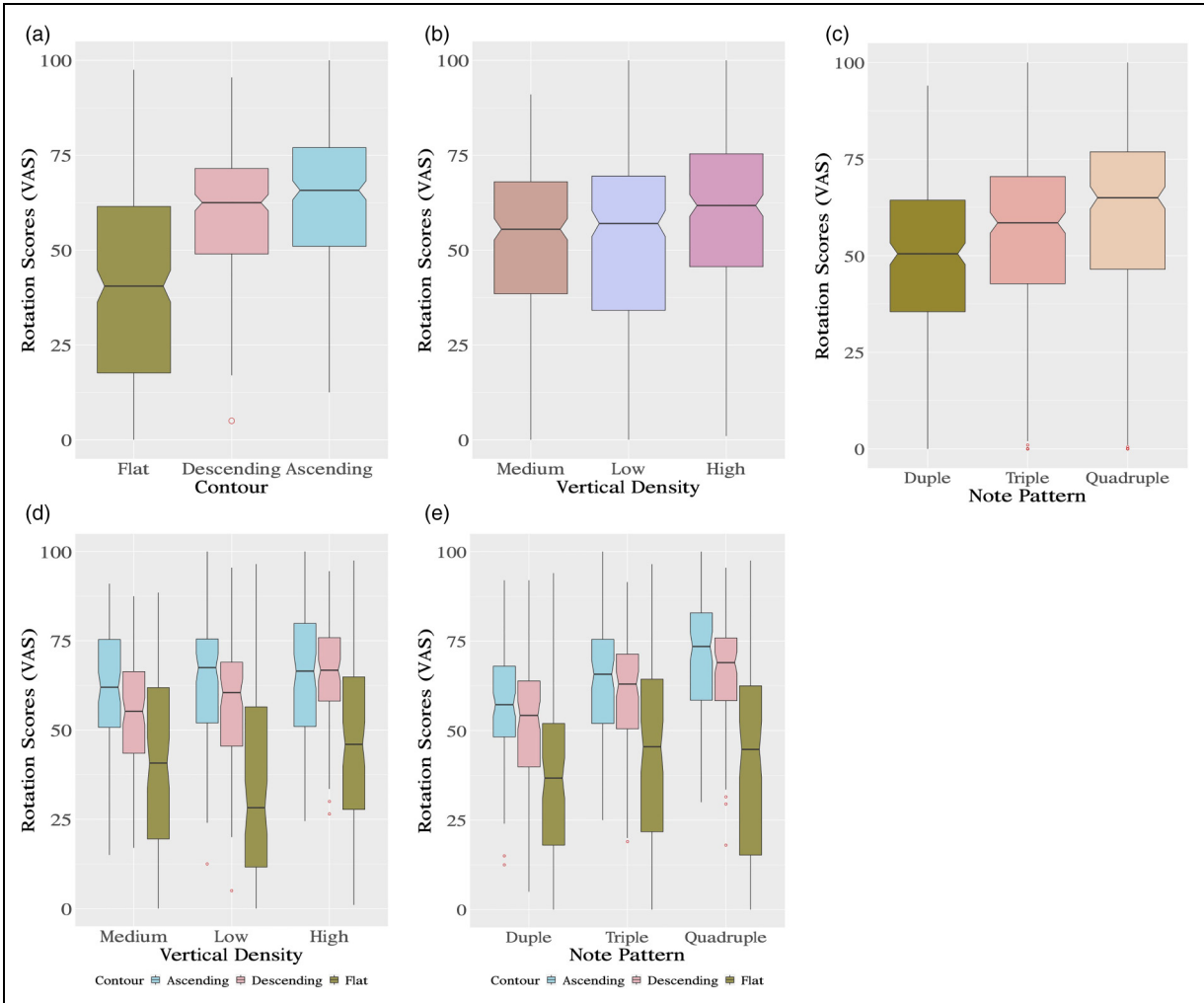
### Physical Involvement

The model explained 42.01% of the variance in Physical Involvement ratings, taking into account the random effects ( $R_m^2 = 0.12$ ,  $R_c^2 = 0.42$ ). The model revealed a significant main effect of Contour ( $\chi^2_{(2)} = 61.05$ ,  $p < .0001$ ), showing that participants felt more Physically Involved in Ascending than with Descending and Flat, and in Descending more than with Flat (Figure 5[a]). A significant main effect for Vertical Density was found ( $\chi^2_{(2)} = 30.93$ ,  $p < .0001$ ), showing that participants felt more Physically Involved with High than with Medium and Low, and with Low less than with Medium (Figure 5[b]). A significant main effect for Note Pattern was found ( $\chi^2_{(2)} = 41.22$ ,  $p < .0001$ ), showing that participants felt more Physically Involved with Quadruple than with Triple and Duple, and with Duple less than with Triple (Figure 5[c]). Significant interaction effects were found for Vertical Density \* Contour ( $\chi^2_{(4)} = 9.77$ ,  $p = .045$ ) (Figure 5[d]; for Tukey's post hoc comparisons for main effects, see Table 5).

### Emotional Involvement

The model explained 53.45% of the variance in Emotional Involvement ratings, taking into account the random effects ( $R_m^2 = 0.11$ ,  $R_c^2 = 0.53$ ). The model revealed a significant main effect of Contour ( $\chi^2_{(2)} = 120.95$ ,  $p < .0001$ ), showing





**Figure 3.** Rotation. Boxplots depicting mean VAS ratings for Rotation with respect to main effects: (a) Contour, (b) Vertical Density, and (c) Note Pattern; and interaction effects for (d) Vertical Density\*Contour, and (e) Note Pattern\*Contour (N = 30) (see text for significant results).

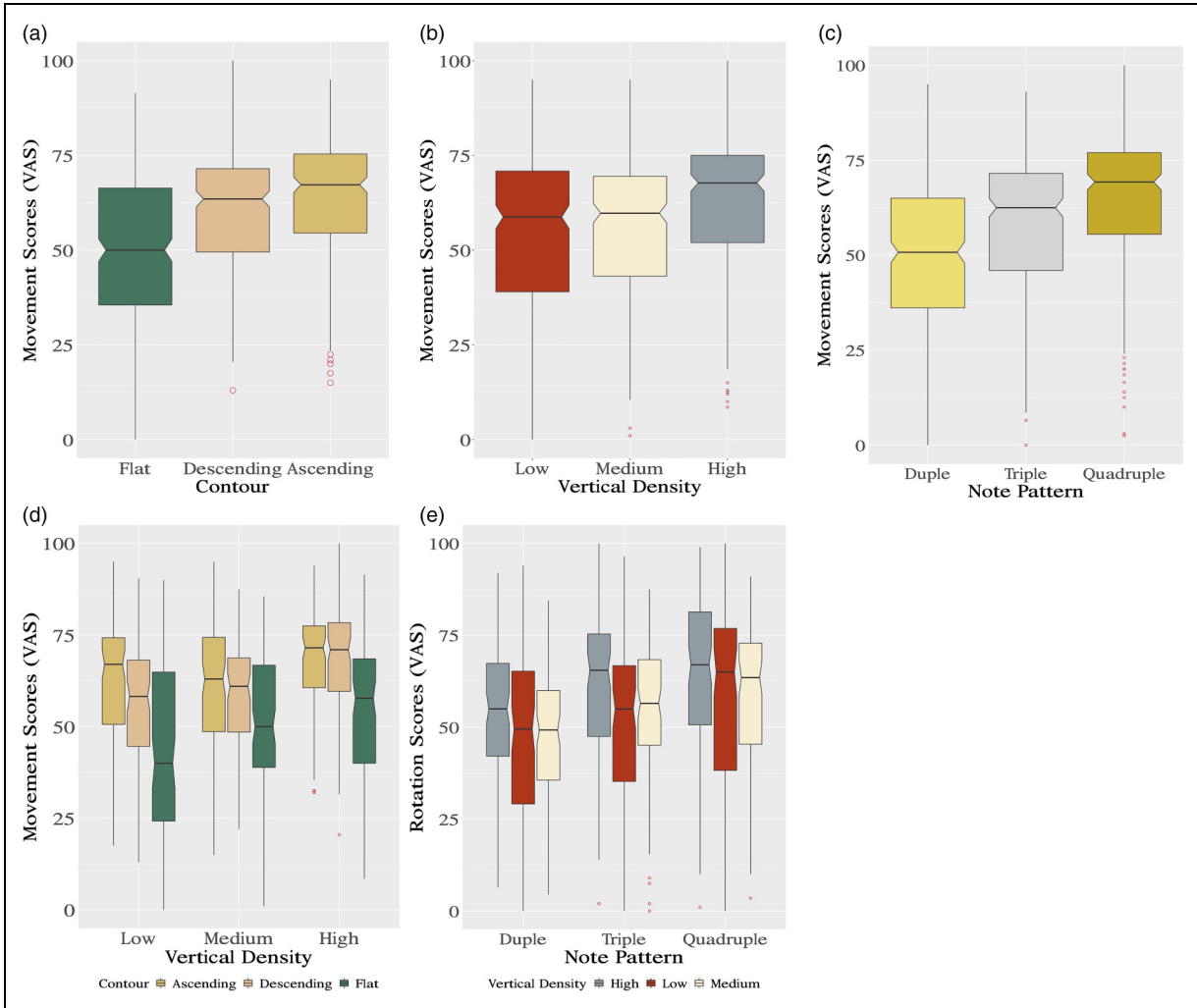
**Table 3.** Tukey’s post hoc comparisons – effect of Contour, Vertical Density, and Note Pattern on perceived rotation.

		Mean Difference	SE	<i>t</i> (754)	<i>p</i>
<i>Contour</i>					
Ascending	Descending	3.99	1.46	2.74	.017 *
	Flat	23.32	1.46	16.02	<.0001 ***
Descending	Flat	19.33	1.46	13.28	<.0001 ***
<i>Vertical Density</i>					
High	Medium	6.56	1.46	4.50	<.0001 ***
	Low	7.92	1.46	5.44	<.0001 ***
Medium	Low	1.36	1.46	.94	.618
<i>Note Pattern</i>					
Quadruple	Triple	3.65	1.46	2.51	.033 *
	Duple	10.54	1.46	7.24	<.0001 ***
Triple	Duple	6.89	1.46	4.73	<.0001 ***

*p* < .05 = \*, *p* < .01 = \*\*, *p* < .001 = \*\*\*

that participants felt more Emotionally Involved with Ascending than with Descending and Flat, and with Descending more than with Flat (Figure 6(a)). A significant main effect for Vertical Density was found ( $\chi^2_{(2)}=22.35$ ,

*p* < .0001), showing that participants felt more Emotionally Involved with High Vertical Density than with Low, and with Low less than with Medium (Figure 6[b]). A significant main effect for Note Pattern



**Figure 4.** Movement. Boxplots depicting mean VAS ratings for Movement with respect to main effects: (a) Contour, (b) Vertical Density, and (c) Note Pattern; and interaction effects for (d) Contour\*Vertical Density, and (e) Note Pattern\*Vertical Density (N = 30) (see text for significant results).

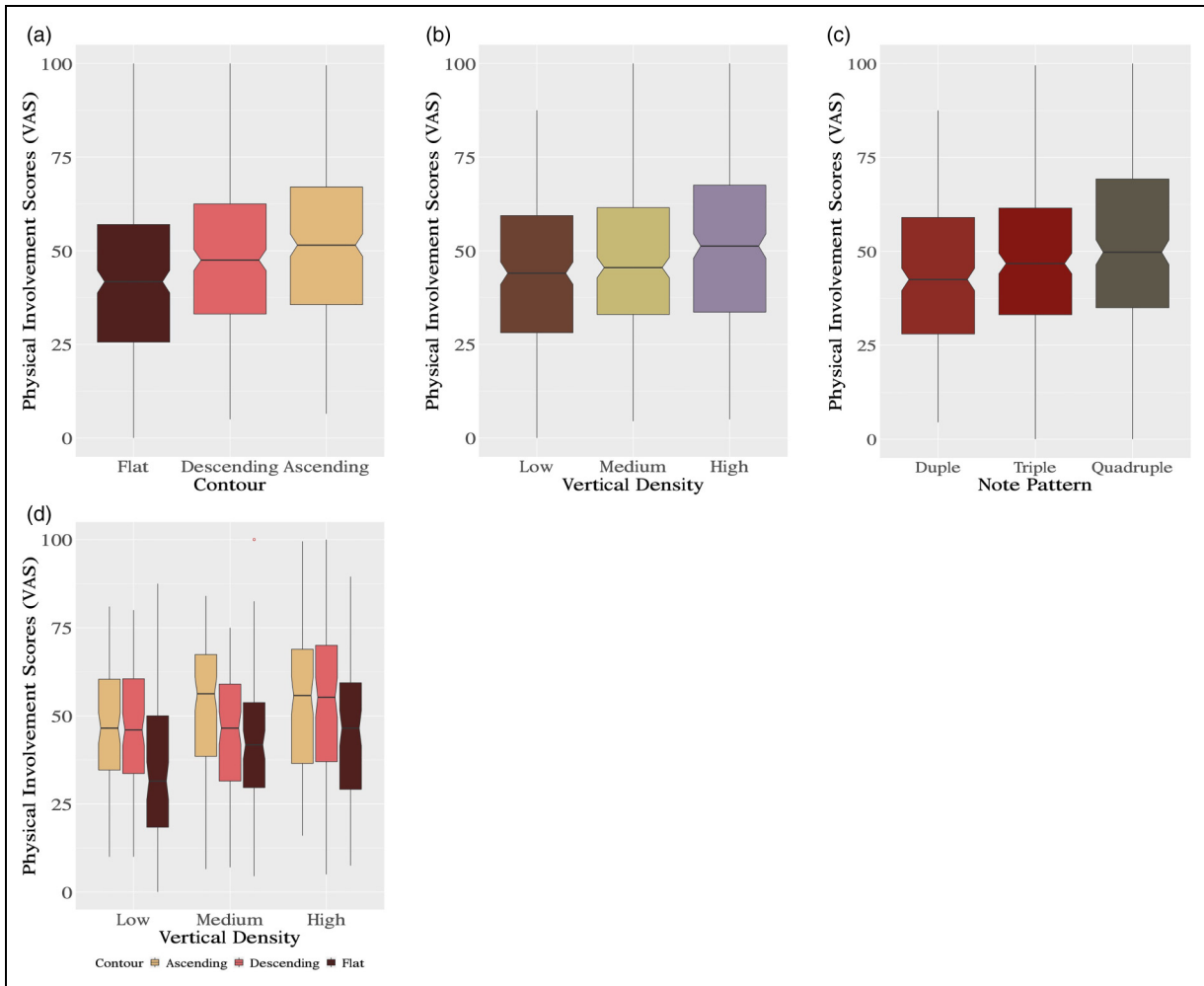
**Table 4.** Tukey’s post hoc comparisons – effect of Contour, Vertical Density, and Note Pattern on perceived movement.

		Mean Difference	SE	<i>t</i> (754)	<i>p</i>
<i>Contour</i>					
Ascending	Descending	4.08	1.30	3.14	.005 **
	Flat	14.96	1.30	11.52	<.0001 ***
Descending	Flat	10.88	1.30	8.38	<.0001 ***
<i>Vertical Density</i>					
High	Medium	7.09	1.30	5.46	<.0001 ***
	Low	9.21	1.30	7.09	<.0001 ***
Medium	Low	2.12	1.30	1.63	.234
<i>Note Pattern</i>					
Quadruple	Triple	6.60	1.30	5.08	<.0001 ***
	Duple	15.12	1.30	11.64	<.0001 ***
Triple	Duple	8.52	1.30	6.56	<.0001 ***

*p* < .05 = \*, *p* < .01 = \*\*, *p* < .001 = \*\*\*

was found ( $\chi^2_{(2)} = 21.30, p < .0001$ ), showing that participants felt less Emotionally Involved with Duple than with

Quadruple, and Triple (Figure 6[c]; for Tukey’s post hoc comparisons for main effects, see Table 6).



**Figure 5.** Physical Involvement. Boxplots depicting mean VAS ratings for Physical Involvement with respect to main effects: (a) Contour, (b) Vertical Density, and (c) Note Pattern; and interaction effects for (d) Vertical Density\*Contour ( $N = 30$ ) (see text for significant results).

**Table 5.** Tukey's post hoc comparisons – effect of Contour, Vertical Density, and Note Pattern on perceived physical involvement.

		Mean Difference	SE	$t_{(754)}$	$p$
<i>Contour</i>					
Ascending	Descending	3.22	1.31	2.46	.04 *
	Flat	10.02	1.31	7.65	<.0001***
Descending	Flat	6.80	1.31	5.19	<.0001***
<i>Vertical Density</i>					
High	Medium	3.97	1.31	3.03	.007 **
	Low	7.27	1.31	5.55	<.0001***
Medium	Low	3.30	1.31	2.52	.03 *
<i>Note Pattern</i>					
Quadruple	Triple	3.75	1.31	2.87	.012 *
	Duple	8.39	1.31	6.41	<.0001***
Triple	Duple	4.64	1.31	3.54	=.001 **

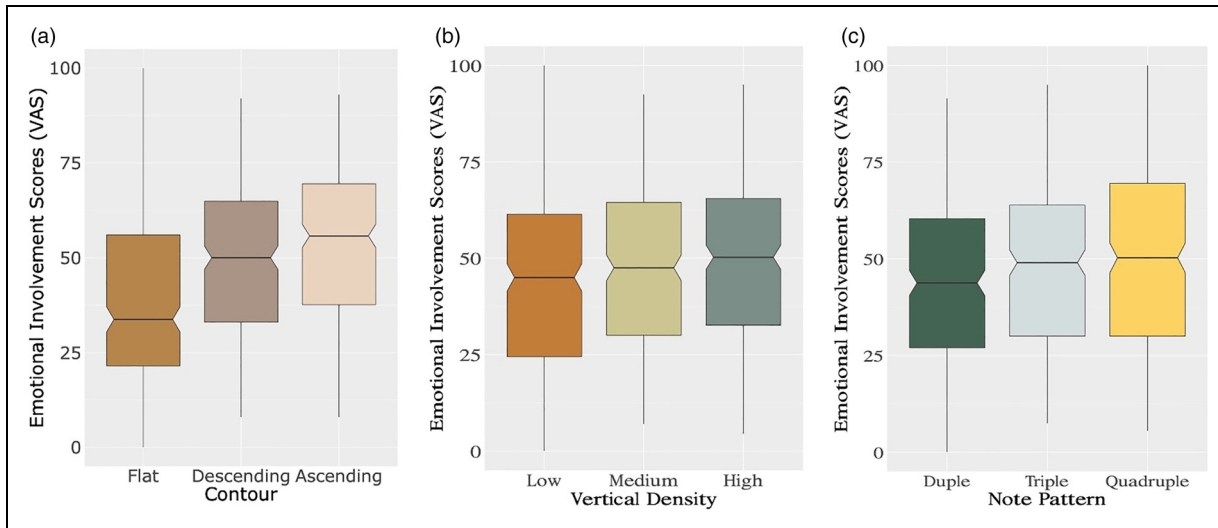
$p < .05 = *$ ,  $p < .01 = **$ ,  $p < .001 = ***$

## Correlations

Results of the Spearman correlations (Table 7, Figure 7) indicate that, after Bonferroni correction ( $p = .05/28 = .002$ ), five positive correlations were significant: Direction \* Physical Involvement ( $R = .64$ ,  $p < .001$ ), Rotation \* Movement ( $R = 0.72$ ,  $p < .0001$ ), Rotation \* Emotional Involvement ( $R = .57$ ,  $p = .001$ ), Movement \* Emotional Involvement ( $R = .63$ ,  $p < .001$ ), and Emotional Involvement \* Physical Involvement ( $R = .59$ ,  $p < .001$ ). No other significant correlations were found between VMIQ-2, BMEQ, and IRI.

## Study I Discussion

In Study 1 we investigated the impact of Contour (Ascending, Descending, Flat), Vertical Density (Low, Medium, High), and Note Pattern (Duple, Triple, Quadruple) on participant ratings of perceived Direction,



**Figure 6.** Emotional Involvement. Boxplots depicting mean VAS ratings for Emotional Involvement with respect to main effects: (a) Contour, (b) Vertical Density, and (c) Note Pattern (N=30); see text for significant results.

**Table 6.** Tukey's post hoc comparisons – effect of Contour, Vertical Density, and Note Pattern on perceived emotional involvement.

		Mean Difference	SE	<i>t</i> (754)	<i>p</i>
<i>Contour</i>					
Ascending	Descending	4.00	1.27	3.16	.005 **
	Flat	13.57	1.27	10.70	<.000 ***
Descending	Flat	9.57	1.27	7.55	<.000 ***
<i>Vertical Density</i>					
High	Medium	1.92	1.27	1.51	.285
	Low	5.88	1.27	4.64	<.000 ***
Medium	Low	3.96	1.27	3.12	.005 **
<i>Note Pattern</i>					
Quadruple	Triple	2.06	1.27	1.62	.237
	Duple	5.77	1.27	4.55	<.000 ***
Triple	Duple	3.72	1.27	2.93	=.01 *

$p < .05 = *$ ,  $p < .01 = **$ ,  $p < .001 = ***$

Movement, Rotation, and Physical and Emotional Involvement in a population of individuals with no professional or semi-professional musical training. Scales to assess Motor Imagination, Trait Empathy, and the tendency to experience Reactive and Affective responses to music were also included to investigate individual differences. In order to extend the ecological validity of the study, a film composer created 27 novel and carefully controlled musical stimuli recorded on a digital piano. The experimental task was to rate the stimuli using a VAS following passive listening. Results of the linear mixed effects analysis show that significant main effects were found for Contour, Vertical Density, and Note Pattern for all dependent variables. Given the high number of Tukey's post hoc comparisons for each interaction, only the most prominent observed trends in results will be discussed. The full list of interactions may be found in the Supplementary Materials 4.

### Perception of Rotation

To test the EToRS, participants were asked about perceived Rotation in the music. It was demonstrated that participants perceive more Rotation in melodies with Ascending Contour than those with Descending Contour, and that they perceive more Rotation in Ascending and Descending than in Flat melodies. This is in line with Hansen and Huron's (2019) result that the sensation of rotation is perceived in loudness/accents patterns consistent with rotating trajectories in pitch that is moving. In line with our hypothesis and Hansen and Huron's (2019) findings, it was demonstrated that Quadruple and Triple Note Patterns were perceived as evoking significantly more Rotation than Duple conditions. However, results also show that Quadruple evokes significantly more Rotation than Triple, contrary to our hypothesis and Hansen and Huron's (2019) findings that Triple Note Patterns were perceived as more spinning/rotating than non-Triple patterns (though the difference is driven primarily

**Table 7.** Spearman's correlations between participants' ratings and questionnaire scores.

		S	R	p
Direction	VMIQ-2 [External]	3461.5	-.06	.779
	VMIQ-2 [Internal]	3593.4	.02	.933
	VMIQ-2	3333.1	-.14	.497
	[Kinesthetic] <sup>a</sup>	2823.0	.14	.492
	BMEQ [Reactive]	2675.8	.40	.027
	Rotation	2886.8	.36	.052
	Movement	1612.7	.64	=.0001 *
	Physical Involvement	2272.8	.49	.005 *
Rotation	Emotional Involvement	3302.1	-.01	.969
	VMIQ-2 [External]			
	VMIQ-2 [Internal]	3709.1	-.02	.939
	VMIQ-2 [Kinesthetic] <sup>a</sup>	2491.4	.15	.470
	BMEQ [Reactive]	2433.3	.26	.195
	Movement	1240.0	.72	<.0001*
	Emotional Involvement	1938.0	.57	=.001*
	Movement	VMIQ-2 [External]	3018.3	.08
VMIQ-2 [Internal]		3271.5	.10	.596
VMIQ-2 [Kinesthetic] <sup>a</sup>		1974.7	.32	.105
BMEQ [Reactive]		2062.1	.37	.057
Physical Involvement		2376.0	.47	.009
Emotional Involvement		1668.0	.63	<.001*
Physical Involvement		VMIQ-2 [External]	3957.8	-.21
	VMIQ-2 [Internal]	3889.3	-.06	.745
	VMIQ-2 [Kinesthetic] <sup>a</sup>	3080.2	-.05	.797
	BMEQ [Reactive]	2527.6	.23	.252
	Rotation	2898.0	.36	.055
	Emotional Involvement	IRI [Empathic Concern]	4280.7	-.17
BMEQ [Affective]		2522.9	.23	.249
Physical Involvement		1844.0	.59	<.001 *

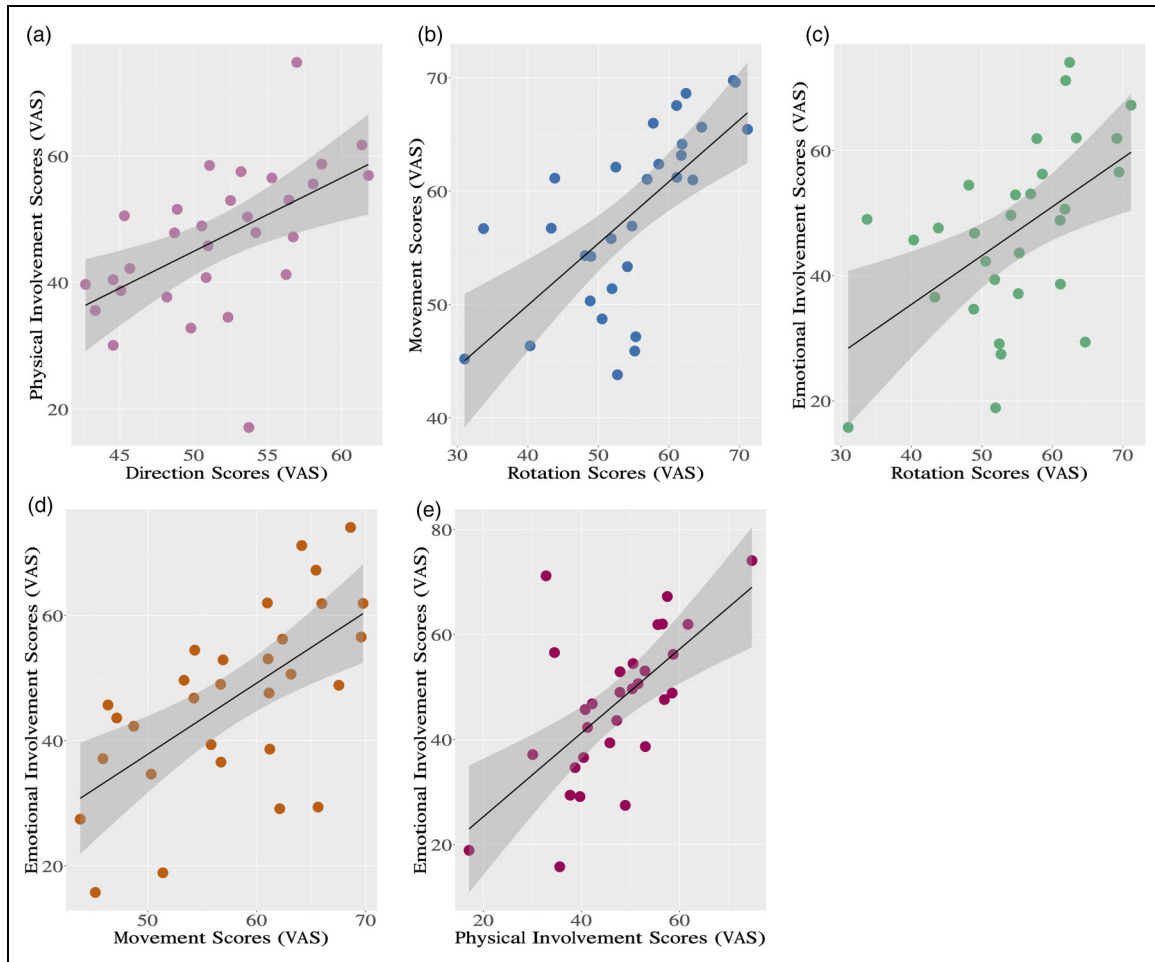
<sup>a</sup>See Supplementary Materials 2.

p < .002 = \*

by low rotating ratings for Duple patterns). As all of our stimuli have a duration of 10 s, and an equal number of bars, Quadruple melodies also have the fastest tempo at the beat level. Indeed, Quadruple melodies have higher note density with four beats per bar, compared with two beats per bar (such as in Duple) or three (such as in Triple). Future studies are needed to determine whether this result may be due to this confound between the melodies' metrical structures and tempo, as Hansen and Huron's (2019) demonstrated findings that perceived Rotation increases with tempo. It was also shown that Vertical Density modulates perceived Rotation, with High Vertical Density evoking greater perceived Rotation than Medium and Low, but no significant differences between Medium and Low. This result may be explained by a common factor between perceived Rotation and perceived Movement.

Significant interactions were observed between Vertical Density \* Contour and Note Pattern \* Contour. No significant differences were observed between Vertical Density conditions for Ascending, while for Descending there were. The High Vertical Density condition adds a higher-pitch melodic line, which increases the average pitch of

the track and may counteract the effect of the Descending Contour. No difference was found between Note Pattern conditions within the Flat Contour condition, only within Ascending and Descending. This may be a threshold effect, as mentioned previously and as consistent with Hansen and Huron (2019), in which the melody needs to be perceived as moving for it to be perceived as rotating. No difference was found between Ascending and Descending within each of the Note Pattern conditions. In all cases, Flat was perceived as evoking the least Rotation, but particularly in the Low Vertical Density condition. It is possible that the combined effects of a Flat Contour and Low Vertical Density lead to the perception of least Movement and thus also least Rotation. Indeed, relationships were observed between participant ratings, finding that the perception of musical Rotation strongly correlates with the perception of musical Movement. This may explain why a similar pattern of interaction effects is observed in the perception of Rotation and in the perception of Movement, as there may be a common factor underlying both of these measures. Many of our participants may have perceived Rotation as just one form of Movement.



**Figure 7.** Correlations. Linear graphs depicting mean VAS ratings for significant correlations ( $N=30$ ) between: (a) Direction and Physical Involvement, (b) Rotation and Movement, (c) Rotation and Emotional Involvement, (d) Movement and Emotional Involvement, and (e) Physical Involvement and Emotional Involvement.

### Pitch Height and Note Density

Results provide moderate support for theories of general (Godøy, 2003; Leman, 2007) and vocal (Cox, 2017) embodied responses to the perception of pitch height and note density, respectively. It was demonstrated that, in line with our hypothesis, Ascending Contour was perceived as moving upward, Descending Contour was perceived as moving downward, and Flat Contour was perceived as moving neither upward nor downward. This result confirms that participants were able to identify the Contour of the melodies, and is consistent with prior literature on a CMC for pitch and vertical motion (Küssner & Leech-Wilkinson, 2014). However, the present study only measured perceived vertical Direction, as is common in the literature (Eitan, 2007), so the possibility of perceived horizontal Direction cannot be discounted, although previous research has suggested pitch Contour to be most strongly associated with vertical movement (Kohn & Eitan, 2009). Language may influence pitch perception, as is well documented in the literature. While verticality is a common means of expressing pitch direction in Western languages (including in Italian,

as the language of this study), other languages may use different metaphors to describe pitch space (e.g., size metaphors for pitch in Croatian [Žuvela & Anić, 2019] and Farsi [Dolscheid et al., 2013]), which should be taken into account if attempting to replicate these results. Results also show that High Vertical Density was perceived to be moving in a higher Direction than Medium and Low, suggesting that the higher melodic line in High Vertical Density tracks enhances the perception of upward movement. This could be due to the High Vertical Density tracks having a higher pitch on average, due to the addition of melodic lines; this will be discussed later.

For the main effect of Note Pattern, in turn, results show that Quadruple is perceived as moving more upward than Triple and Duple, and Triple more so than Duple. The Quadruple Note Pattern was also associated with increased Physical Involvement, possibly due to the higher rate of notes, so it is possible that Physical Involvement may contribute to the perception of Direction in the music. This could be explained in terms of embodied vocal responses that link pitch height and effort (Cox, 2017), as it was

found that participant ratings for perceived Direction correlate with Physical Involvement, although causality cannot be assumed.

For the Movement measure, results show that, even though the tempo, interval size, and note density were equivalent in all conditions, participants not only perceived more Movement in Ascending and Descending conditions with respect to Flat, but also more Movement in Ascending than in Descending. These results provide preliminary evidence for the hypothesis that Contour modulates perceived Movement in participants. Eitan and Granot (2006) also found rises in pitch to be associated with increased intensity, consistent with embodied vocal responses that link pitch height and effort (Cox, 2017). However, contrary to our results, they found a stronger effect of perceived motion in Descending rather than Ascending pitch Contour. Importantly, it was also shown that High Vertical Density melodies are perceived as evoking greater perceived Movement than Medium and Low, and Medium more so than Low. These results suggest that participants may associate greater Vertical Density with a greater number of sound-producing gestures needed to produce the musical events, and thus that notes are indeed perceived as kinetic events, in line with the notion of motormimetic elements in music perception (Cox, 2017; Godøy, 2003; Godøy & Leman, 2010). Our results also demonstrate that Quadruple is perceived as evoking greater Movement than Triple and Duple, and Triple more than Duple, indicating again that note density and/or tempo modulate participants' perception of Movement.

However, a significant Note Pattern \* Vertical Density interaction indicates that there is a threshold effect for the perception of Movement in Quadruple melodies of different Vertical Density (i.e., no significant differences between Quadruple Low, Quadruple Medium and Quadruple High). The addition of more melodic lines seems to have no effect within the Quadruple condition, which was already rated very high overall for Movement, suggesting that there may be a ceiling effect of the Quadruple condition. Since the Quadruple condition already has a high note density, increasing Vertical Density leads to a more pronounced "co-articulation" effect, where there is a "smearing" of individual sounds so that they are no longer perceived as separate events (Godøy, 2008).

The effect of Contour on perceived Movement was consistent between Note Pattern conditions. An effect was found between Ascending and Descending but only in the Low Vertical Density condition, possibly because the Contour was more salient in Low Vertical Density. In effect, it appears that increasing Vertical Density reduces the difference between Ascending and Descending conditions, as there is no difference between Ascending and Descending after the addition of the lower pedal notes in Medium Vertical Density. We have no theoretical justification for this finding but future research may investigate it.

As for the Physical Involvement measure, results demonstrate that Ascending Contour is perceived as more Physically Involving than both Descending and Flat

Contour, and Descending more than Flat. This follows the same pattern of results as for perceived Movement, showing that participants associate Ascending Contour with greater exertion or effort, in line with the "greater [exertion] is higher" embodied metaphor (Cox, 2017). Participants also perceived High Vertical Density melodies to be more Physically Involving than Low and Medium Vertical Density, and Medium more than Low, providing further support for theories of general embodied responses to music (e.g., Godøy, 2003). For Note Pattern, in turn, Quadruple is perceived as more Physically Involving than Triple and Duple, and Triple more than Duple, providing further evidence in support of the modulating effect of note density on Physical Involvement. Indeed, this self-report measure may indicate a tacit desire to move. For instance, Ascending pitch may induce more energetic movement (Kohn & Eitan, 2009; Küssner et al. 2014), consistent with self-reported Physical Involvement.

An interaction was also observed between Vertical Density and Contour. The addition of the pedal notes in Medium Vertical Density increased Physical Involvement in the Ascending Condition, while the addition of the higher melodic line in High Vertical Density increased Physical Involvement in Descending. There may be two separate processes behind this. Bass notes often induce more Movement during music listening (Burger et al., 2012), which may have a more pronounced effect on the Ascending melody as they are more distinguished from the high endpoint of the melody. Alternatively, the pedal note in the Medium Vertical Density condition may also serve as a reference point, increasing the salience of the melodic Contour. Conversely, the higher pitch introduced in High Vertical Density may increase feelings of tension due to vocal embodied responses (Cox, 2017), specifically in the Descending Contour, which otherwise could be perceived as having decreasing tension as the pitch decreases. This may have nullified the effects of Contour within the High conditions.

### *Music, Movement, and Emotional Involvement*

The induction of emotion from music, also known as emotional contagion, may have an embodied component, as suggested by the SAME theory (Overy & Molnar-Szakacs, 2009). We tested this with a general Emotional Involvement measure, and found significant main effects as well as an association with Physical Involvement. Participants perceive melodies with Ascending Contour to be more Emotionally Involving than both Flat and Descending melodies, and Descending more so than Flat. This result suggests that the stronger musical affect evoked by Ascending melodies may be driven by greater perceived exertion, providing preliminary support for embodied vocal response effects (Cox, 2017) and potentially connecting it with the SAME theory (Overy & Molnar-Szakacs, 2009), as the increased vocal tension may contribute to music-induced emotion. Indeed, melodic high-points regularly coincide with climactic moments in a piece of music

(Eitan, 1997). Previous research has also found an association between volume and vertical motion, with increases in loudness being perceived as moving upward (Eitan et al., 2008). The means for the loudness analysis do differ slightly between Ascending, Descending, and Flat, although the difference between Ascending and Descending is not consistent between objective measures. Results also demonstrate that High Vertical Density is more Emotionally Involving than Low, and Medium more so than Low, suggesting that vertical note density may modulate Emotional Involvement, but the difference in note density between High and Medium Vertical Density is not salient enough for such a modulation. The results for Note Pattern, with Quadruple perceived as being more Emotionally Involving than Duple, and Triple more than Duple, may indicate a similar modulating effect of horizontal note density.

A relationship was also observed between Emotional Involvement and Physical Involvement, with a trend toward a correlation between Physical Involvement and perceived Movement. This provides evidence in support of the theory of the co-representation of musical experience (i.e., SAME theory), which posits that musical meaning is conveyed through implied Movement, and may be pivotal for musical empathy (Molnar-Szakacs & Overy, 2006; Overy & Molnar-Szakacs, 2009). There was no correlation between the IRI (Davis, 1980) and Emotional or Physical Involvement. This may have been expected, as previous research has found Empathy to correlate with embodied and motor responses to music (Bamford & Davidson, 2019; Wallmark et al., 2018). The results here are only correlational, thus it is not clear whether perceived Movement may cause Emotional Involvement as the SAME theory would claim. Similarly, no correlations were found between the BMEQ, VMIQ-2, and the involvement measures, although this was purely exploratory, as the VMIQ-2 has not been previously used in music listening tasks.

## STUDY 2

Study 1 investigated embodied CMCs for a novel set of controlled musical stimuli. However, this study identified some potential confounds with basic musical features within the stimuli. First, it was apparent that the effects of Vertical Density could be due to a confound with perceived pitch height. The stimuli marked as Low Vertical Density in Study 1 had a single melodic line, while Medium Vertical Density added a pedal note for additional harmony. Meanwhile, High Vertical Density added one or more additional melodic lines. In this study, Vertical Density then refers to the number of harmonic voices, but there is a qualitative difference between having multiple melodic voices (as in the High condition) compared with having one melodic voice and a pedal note (as in the Medium condition). Furthermore, the High Vertical Density condition is usually higher in pitch than the other two, due to space between the voices. Therefore, the results for Vertical Density may be due either to the increased number of

melodic voices, or to the higher average pitch of the High Vertical Density condition relative to the others.

Given that there is a common perceptual association between Pitch and Loudness (Küssner et al., 2014), perceived loudness could also have been a factor, both for Vertical Density and for Note Pattern. Because tempo at the bar level was held constant across all stimuli, Note Pattern could have been confounded with Speed at the beat level, with the Quadruple condition having the highest number of quavers per bar, followed by Triple, then Duple.

Given these potential perceptual confounds, we conducted an additional study in which a new sample of participants was presented with the same stimuli from Study 1 but with a different set of questions. Specifically, Study 2 asked participants to rate each stimulus for its Pitch, Speed, and Loudness. Specifically, we expected: 1) High Vertical Density to be associated with higher Pitch, as well as higher Loudness and higher Speed; 2) Quadruple Note Pattern to be associated with higher Speed, as well as higher Loudness and Pitch; and 3) Pitch ratings to correlate with those of Loudness and Speed. Other possible associations were tested for exploratory purposes in order to better describe how the stimuli may be perceived.

## Materials and Methods

### Participants

Participants were recruited with the same criteria as in Study 1, except that residents from all over Italy were included in Study 2. The average total number of years of musical training was  $1.57 \pm 1.75$  SD years. In total, 56 healthy volunteers of Italian nationality took part in the study: 29 females and 27 males, mean age 28.34 ( $SD = 6.03$ ,  $min = 18$ ,  $max = 35$ ). Power was calculated a priori by means of G\*Power 3.1 (Faul et al., 2007) using the linear multiple regression: random model to test for a linear mixed effect model for each dependent variable. With a Cohen's F effect size equal to 0.25 (medium effect size), an alpha level of .05, 3 predictors, and a power of 0.9 resulted in a total sample size of 51. All participants provided written informed consent to participate in the study, which was conducted in accordance with the Declaration of Helsinki (Ndebele, 2013), complying with the ethical standards of the Italian Board of Psychologists as well as the Ethical Code for Psychological Research of the Italian Psychological Society and the Ethical Committee of the Area Vasta Emilia Nord (AVEN).

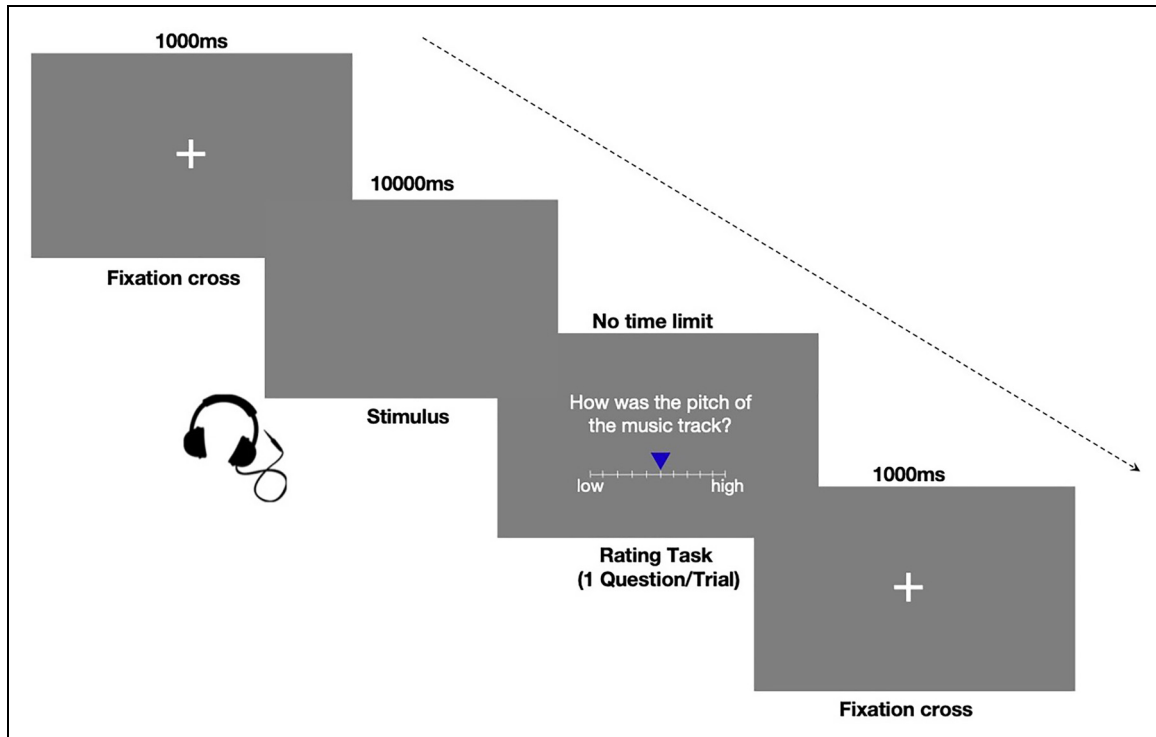
### Stimuli Creation

The same stimulus set as in Study 1 was used.

### Procedure

Participants carried out Study 2 using Pavlovia (the online version of PsychoPy 3.0 software) on their laptops or





**Figure 8.** Example of Study 2 experimental trial. Components: fixation cross frame (1,000 ms), stimulus frame (10,000 ms) and rating task (no time limit). Experiment was created using Pavlovia 3.0.

computers at home. In order to provide support to the participant and monitor their progress, the experimenter and participant maintained contact during the experimental procedure via audio call using mobile phones. The experimenter and participant kept their video and microphone on during the instructions phase. When participants began the experiment, both experimenter and participant turned off their video and microphone in order to ensure privacy and avoid conditioning. Prior to beginning the experiment, participants were instructed to ensure a stable internet connection on their laptop/computer and mobile phone, a quiet environment, and to use earphones. The experimental task run by Pavlovia consisted of 27 audio tracks presented in random order. In each trial, a fixation cross was presented for 1,000 ms, an audio stimulus was presented for 10,000 ms (with a grey screen), and afterward a question was presented with no time limit (Figure 8). Each stimulus was followed by one question at a time, in randomized order, asking the participant to rate the stimulus on one of the following dimensions: 1) Pitch (low–high), 2) Loudness (soft–loud), and 3) Speed (slow–fast) (for the original Italian and English translations, see Supplementary Materials 5). Specifically, “loud” is denoted as “forte,” and “soft” is denoted as “piano” (“forte,” 2010), which in Italian refer to the effort needed to produce a given musical sound rather than a quality of the music itself. In addition, “forte” may also refer to emotional “intensity” or “coolness,” as reported by participants in the pilot. To avoid confusion, we phrased the question as “How did the musical track seem to be played?” (*Come ti sembrava suonato il*

*brano musicale?*), with a scale ranging from “softly” (*piano*) to “strongly” (*forte*). Participants were asked to listen to the stimuli and answer the questions as quickly and as accurately as possible (no specific time limit was given), using the mouse to move a blue cursor (positioned in the middle) along a VAS that recorded responses on a scale ranging from 0 to 100 (numerical values were not visible, only the labels). The study design was analogous to that of Study 1: 3 Contours (Ascending, Descending, Flat) \* 3 Note Patterns (Duple, Triple, Quadruple) \* 3 Vertical Densities (High, Medium, Low), for a total of 27 conditions. Each experimental condition was repeated six times (three questions and two repetitions), for a total of 162 presented trials. Before carrying out the experimental procedure participants were presented with a brief training phase to become accustomed with the task. After the experimental session participants were asked to fill out a short debriefing survey about their experience.

### Analysis

Data were analyzed in the same manner as in Study 1, with different dependent variables. A linear mixed effects analysis was carried out for each dependent variable (perceived Pitch, Speed, and Loudness). The same independent fixed variables were used as in Study 1 (Contour, Vertical Density, and Note Pattern), and participant intercepts were included as random effects. Post hoc comparisons were done using Tukey’s tests with Bonferroni correction for multiple comparisons. All analyses were performed

using R software (R Core Team, 2019) and lme4 (Bates et al., 2015), ordinal (Christensen, 2019), effects (Fox, 2003), and emmeans (Lenth et al., 2020) functions; for data visualization, the ggplot2 package was used (Wickham, 2016). Audio analysis was completed using the MIR Toolbox (Lartillot & Toiviainen, 2007).

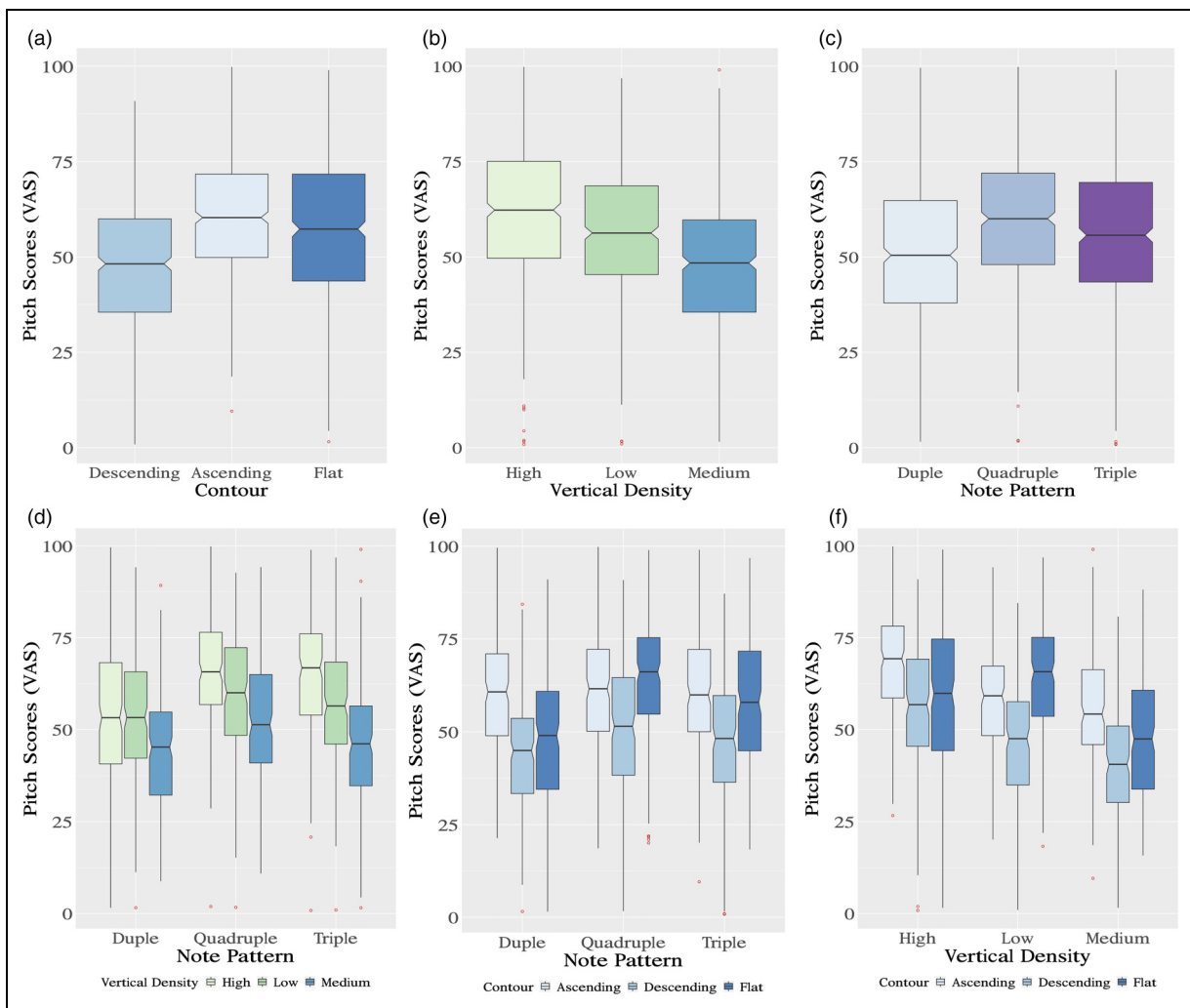
## Results

Estimated marginal means calculated from the linear mixed effects model for all conditions, as well as test statistics for all Tukey's post hoc comparisons for interaction effects, may be found in Supplementary Materials 6 and 7, respectively.

### Pitch

The model explained 42.68% of the variance in Pitch ratings, taking into account the random effects ( $R_m^2 = 0.28$ ,  $R_c^2 = 0.43$ ). The model revealed a significant main

effect of Contour ( $\chi^2_{(2)} = 233.76$ ,  $p < .0001$ ), showing that participants perceived Ascending Contour to have a higher Pitch than Descending and Flat, and Flat to have a higher Pitch than Descending (Figure 9[a]). A significant main effect for Vertical Density was found ( $\chi^2_{(2)} = 233.62$ ,  $p < .0001$ ), showing that participants perceived High Vertical Density melodies to have a higher Pitch than Medium and Low, and Low to have a higher Pitch than Medium (Figure 9[b]). A significant main effect for Note Pattern was found ( $\chi^2_{(2)} = 81.34$ ,  $p < .0001$ ), showing that participants perceived Quadruple to have a higher Pitch than Triple and Duple, and Triple to have a higher Pitch than Duple (Figure 9[c]). Significant interaction effects were found for Note Pattern \* Vertical Density ( $\chi^2_{(4)} = 25.78$ ,  $p < .0001$ ) (Figure 9[d]), Note Pattern \* Contour ( $\chi^2_{(4)} = 25.78$ ,  $p < .0001$ ) (Figure 9[e]), Vertical Density \* Contour ( $\chi^2_{(4)} = 79.71$ ,  $p < .0001$ ) (Figure 9[f]), and Note Pattern \* Contour \* Vertical Density ( $\chi^2_{(8)} = 59.30$ ,  $p < .0001$ ) (for Tukey's post hoc comparisons for main effects, see Table 8).



**Figure 9.** Pitch. Boxplots depicting mean VAS ratings for Pitch with respect to main effects: (a) Contour, (b) Vertical Density, and (c) Note Pattern; and interactions between (d) Note Pattern\*Vertical Density, (e) Note Pattern\*Contour, and (f) Vertical Density\*Contour (N = 56) (see text for significant results).

**Table 8.** Tukey's post hoc comparisons – effect of Contour, Vertical Density, and Note Pattern on perceived pitch.

		Mean Difference	SE	$t_{(1426)}$	$p$
<i>Contour</i>					
Ascending	Descending	12.80	0.88	14.51	<.0001***
	Flat	3.68	0.88	4.17	=.0001***
Descending	Flat	-9.11	0.88	-10.34	<.0001***
<i>Vertical Density</i>					
High	Medium	13.35	0.88	15.13	<.0001***
	Low	5.04	0.88	5.71	<.0001***
Medium	Low	-8.31	0.88	-9.43	<.0001***
<i>Note Pattern</i>					
Quadruple	Triple	3.62	0.88	4.10	=.0001***
	Duple	7.96	0.88	9.02	<.0001***
Triple	Duple	4.34	0.88	4.92	<.0001***

$p < .05 = *$ ,  $p < .01 = **$ ,  $p < .001 = ***$

### Speed

The model explained 65.34% of the variance in Speed ratings, taking into account the random effects ( $R_m^2 = 0.46$ ,  $R_c^2 = 0.65$ ). The model revealed a significant main effect of Contour ( $\chi^2_{(2)} = 204.99$ ,  $p < .0001$ ), showing that participants perceived Ascending Contour to have a greater Speed with respect to Descending, and Flat to have a greater Speed with respect to Ascending and Descending (Figure 10[a]). A significant main effect for Vertical Density was found ( $\chi^2_{(2)} = 122.27$ ,  $p < .0001$ ), showing that participants perceived High Vertical Density melodies to have a greater Speed than Medium and Low, and Low to have a greater Speed than Medium (Figure 10[b]). A significant main effect for Note Pattern was found ( $\chi^2_{(2)} = 1509.35$ ,  $p < .0001$ ), showing that participants perceived Quadruple to have a greater Speed than Triple and Duple, and Triple to have a greater Speed than Duple (Figure 10[c]). Significant interaction effects were found for Note Pattern \* Vertical Density ( $\chi^2_{(4)} = 27.92$ ,  $p < .0001$ ) (Figure 10[d]), and for Contour \* Vertical Density ( $\chi^2_{(4)} = 115.03$ ,  $p < .0001$ ) (Figure 10[a]; for Tukey's post hoc comparisons for main effects, see Table 9).

### Loudness

The model explained 40.44% of the variance in Loudness ratings, taking into account the random effects ( $R_m^2 = 0.23$ ,  $R_c^2 = 0.40$ ). The model revealed a significant main effect of Contour ( $\chi^2_{(2)} = 50.93$ ,  $p < .0001$ ), showing that participants perceived Ascending Contour to evoke less Loudness than Flat, and Descending to evoke less Loudness than Flat (Figure 11[a]). A significant main effect for Vertical Density was found ( $\chi^2_{(2)} = 145.15$ ,  $p < .0001$ ), showing that participants perceived High Vertical Density melodies to evoke greater Loudness than Medium and Low, and with Medium evoking greater Loudness with respect to Low (Figure 11[b]). A significant main effect for Note Pattern was found ( $\chi^2_{(2)} = 344.76$ ,  $p <$

.0001), showing that participants perceived Quadruple to evoke greater Loudness than Triple and Duple, and Triple more than Duple (Figure 11[c]). Significant interaction effects were found for Note Pattern \* Vertical Density ( $\chi^2_{(4)} = 10.60$ ,  $p = .031$ ) (Figure 11[d]), for Vertical Density \* Contour ( $\chi^2_{(4)} = 23.00$ ,  $p < .001$ ) (Figure 11[e]), and for Note Pattern \* Vertical Density \* Contour ( $\chi^2_{(8)} = 17.15$ ,  $p = .03$ ) (for Tukey's post hoc comparisons for main effects, see Table 10).

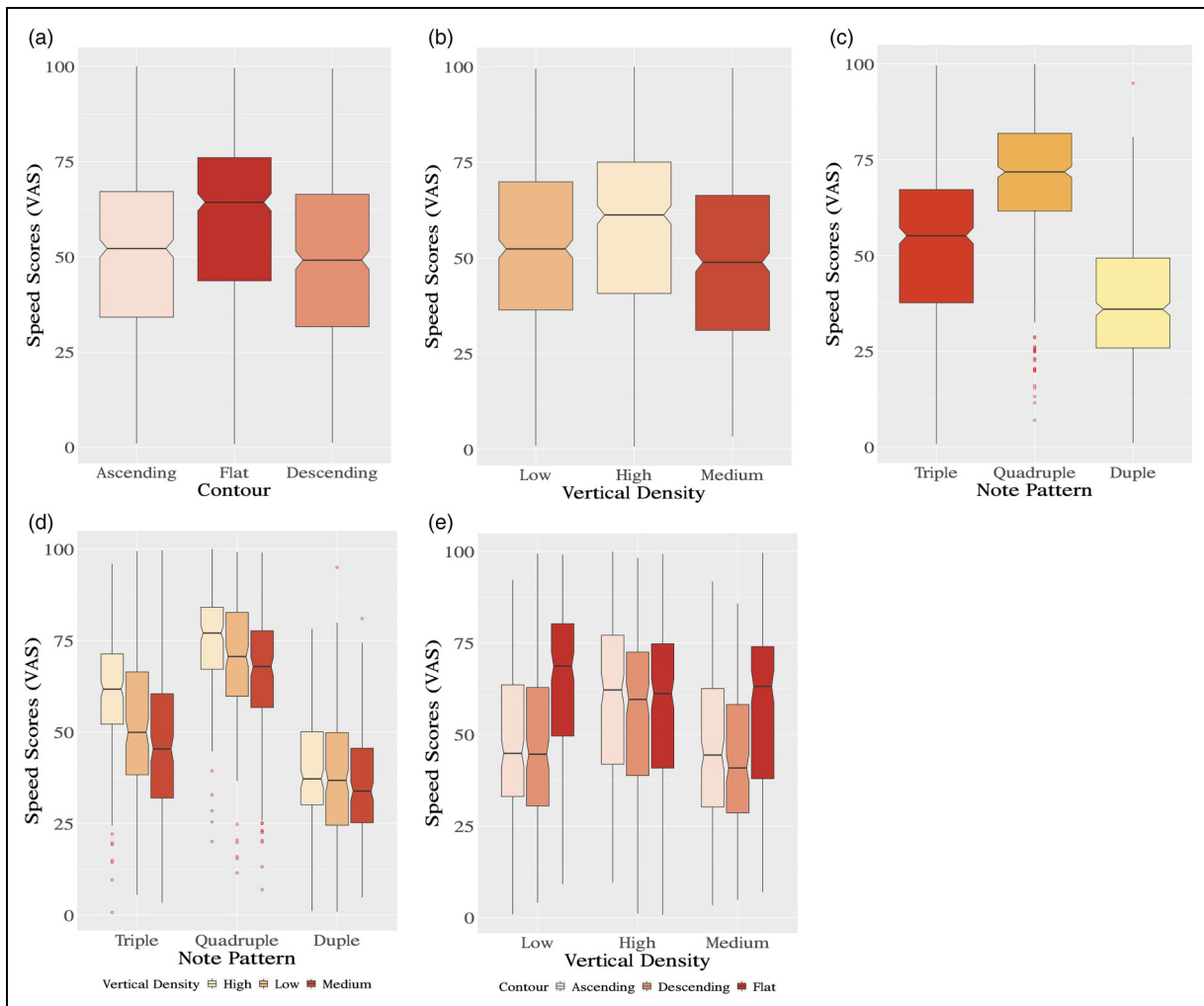
An objective measure of loudness was calculated using the Root-Mean-Square Energy (RMS) and Loudness Units to Full Scale (LUFS) of each of the musical stimuli in the MIR Toolbox (Lartillot & Toivainen, 2007), and the Equivalent Continuous Sound Level (LAeq) in decibels, using the splMeter function built into MATLAB. Descriptive statistics of these are presented in Table 11. No statistical tests were done, as these descriptive statistics describe the entire stimulus set ( $N = 27$ ), grouped by condition. Results for RMS and LUFS found that for Contour, Flat was Louder than Descending, and Descending Louder than Ascending; for Note Pattern, Quadruple was Louder than Triple, and Triple Louder than Duple; and for Vertical Density, High was Louder than Medium, and Medium was Louder than Low. The LAeq found the same, except for Contour, which may be due to the frequency weighting used in LAeq analysis.

### Correlations

Results of the Spearman correlations (Table 12, Figure 12) indicate that after Bonferroni correction ( $p = .05/3 = .017$ ), three positive correlations resulted significant: Loudness \* Pitch ( $R = 0.48$ ,  $p < .001$ ), Loudness \* Speed ( $R = 0.55$ ,  $p < .0001$ ), and Pitch \* Speed ( $R = 0.41$ ,  $p = .002$ ).

### Study 2 Discussion

In Study 2, we investigated the impact of Contour (Ascending, Descending, Flat), Vertical Density (Low,



**Figure 10.** Speed. Boxplots depicting mean VAS ratings for Speed with respect to main effects: (a) Contour, (b) Vertical Density, and (c) Note Pattern; and interactions between (d) Note Pattern\*Vertical Density, and (e) Vertical Density\*Contour (N = 56) (see text for significant results).

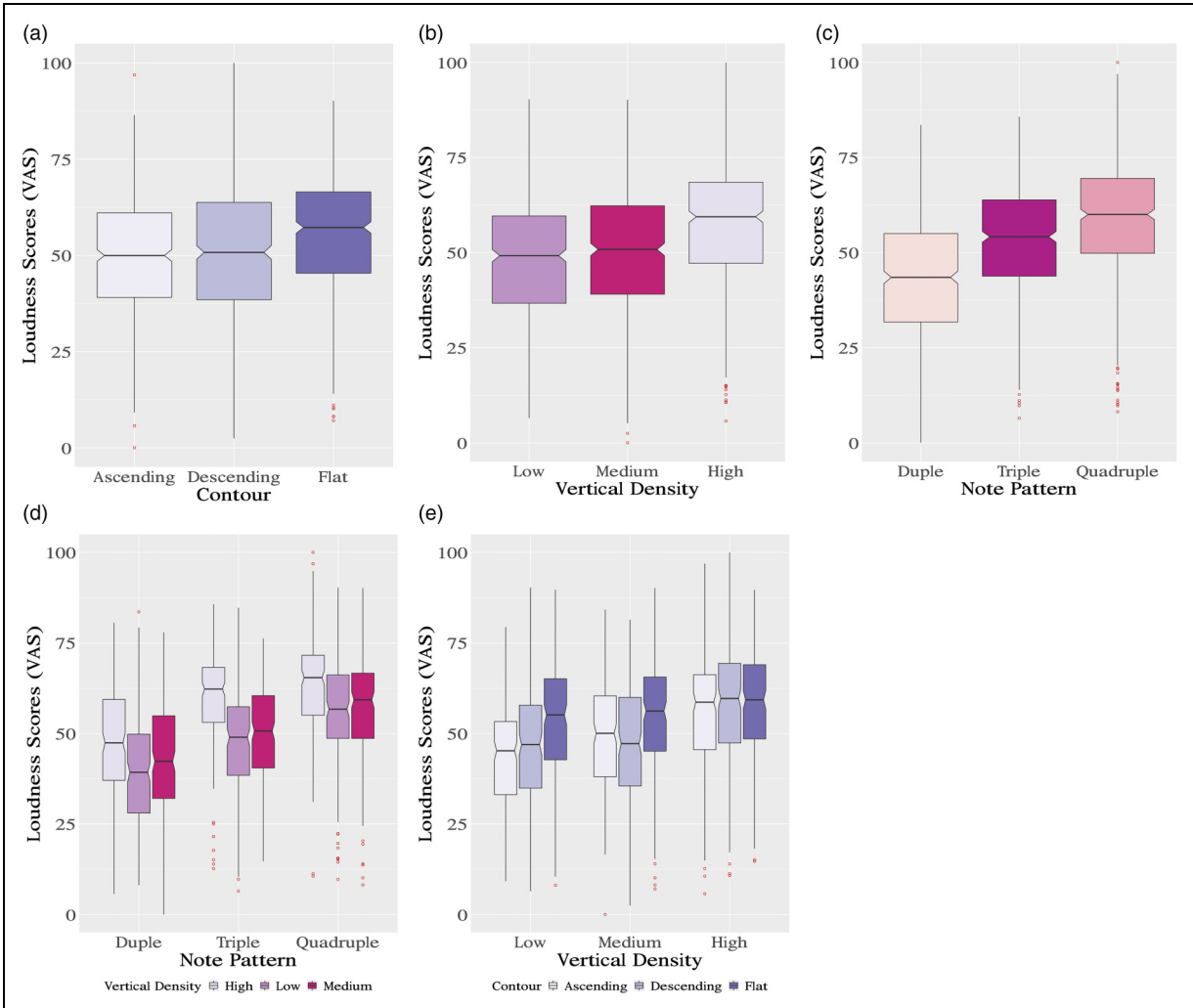
**Table 9.** Tukey's post hoc comparisons – effect of Contour, Vertical Density, and Note Pattern on perceived speed.

		Mean Difference	SE	$t_{(1429)}$	$p$
<i>Contour</i>					
Ascending	Descending	2.16	0.82	2.64	.02*
	Flat	-8.89	0.82	-10.87	<.000 ***
Descending	Flat	-11.06	0.82	-13.51	<.000 ***
<i>Vertical Density</i>					
High	Medium	8.94	0.82	10.93	<.000 ***
	Low	5.64	0.82	6.89	<.000 ***
Medium	Low	-3.30	0.82	-4.04	<.001 ***
<i>Note Pattern</i>					
Quadruple	Triple	17.7	0.82	21.61	<.000 ***
	Duple	31.7	0.82	38.76	<.000 ***
Triple	Duple	14.0	0.82	17.16	<.000 ***

$p < .05 = *$ ,  $p < .01 = **$ ,  $p < .001 = ***$

Medium, High), and Note Pattern (Duple, Triple, Quadruple) on participant ratings of perceived Pitch, Loudness, and Speed in a population of individuals with

no professional or semi-professional musical training. The same set of stimuli was used as in Study 1 (27 novel digital piano tracks). The experimental task, conducted



**Figure 11.** Loudness. Boxplots depicting mean VAS ratings for Loudness with respect to main effects: (a) Contour, (b) Vertical Density, and (c) Note Pattern; and interactions between (d) Note Pattern\*Vertical Density, and (e) Vertical Density\*Contour (N= 56) (see text for significant results).

**Table 10.** Tukey’s post hoc comparisons – effect of Contour, Vertical Density, and Note Pattern on perceived loudness.

		Mean Difference	SE	t (1430)	p
<i>Contour</i>					
Ascending	Descending	-0.77	0.82	-0.94	.62
	Flat	-5.38	0.82	-6.60	<.0001***
Descending	Flat	-4.61	0.82	-5.66	<.0001***
<i>Vertical Density</i>					
High	Medium	7.16	0.82	8.78	<.0001***
	Low	9.40	0.82	11.53	<.0001***
Medium	Low	2.24	0.82	2.75	.017*
<i>Note Pattern</i>					
Quadruple	Triple	6.01	0.82	7.37	<.0001***
	Duple	15.04	0.82	18.44	<.0001***
Triple	Duple	9.03	0.82	11.08	<.0001***

p < .05 = \*, p < .01 = \*\*, p < .001 = \*\*\*

online, was to rate the stimuli using a VAS following passive listening. Results of the linear mixed effects analysis show that significant main effects were found for Contour, Vertical Density, and Note Pattern for all

dependent variables. Given the very high number of Tukey’s post hoc comparisons for each interaction, only the most prominent observed trends in the results of the two-way interactions will be discussed, and three-way

**Table 11.** Descriptive statistics of the RMS energy, LUFS and LAeq for each stimulus condition.

		RMS Mean	SD	LUFS Mean	SD	LAeq Mean	SD
Contour	Ascending	0.0320	0.0042	-24.8	1.37	63.6	1.23
	Descending	0.0331	0.0073	-24.8	2.03	62.8	2.43
	Flat	0.0336	0.0056	-23.9	1.34	64.1	1.31
Note Pattern	Duple	0.0307	0.0058	-25.2	1.82	62.6	2.04
	Quadruple	0.0348	0.0046	-23.9	1.17	64.0	1.34
	Triple	0.0331	0.0062	-24.3	1.66	63.8	1.71
Vertical Density	High	0.0382	0.0037	-23.1	0.836	64.7	1.33
	Low	0.0270	0.0035	-25.8	1.45	62.3	1.53
	Medium	0.0328	0.0035	-24.6	1.20	63.5	1.70

**Table 12.** Spearman's correlations between participants' ratings.

		S	R	
Pitch	Speed	17,338	0.41	.002*
Speed	Loudness	15,144	0.48	<.001*
	Loudness	13,310	0.55	<.0001*

$p < .017 = *$

interactions will not be discussed. The full list of two-way interactions may be found in the Supplementary Materials 7.

### Pitch and Loudness

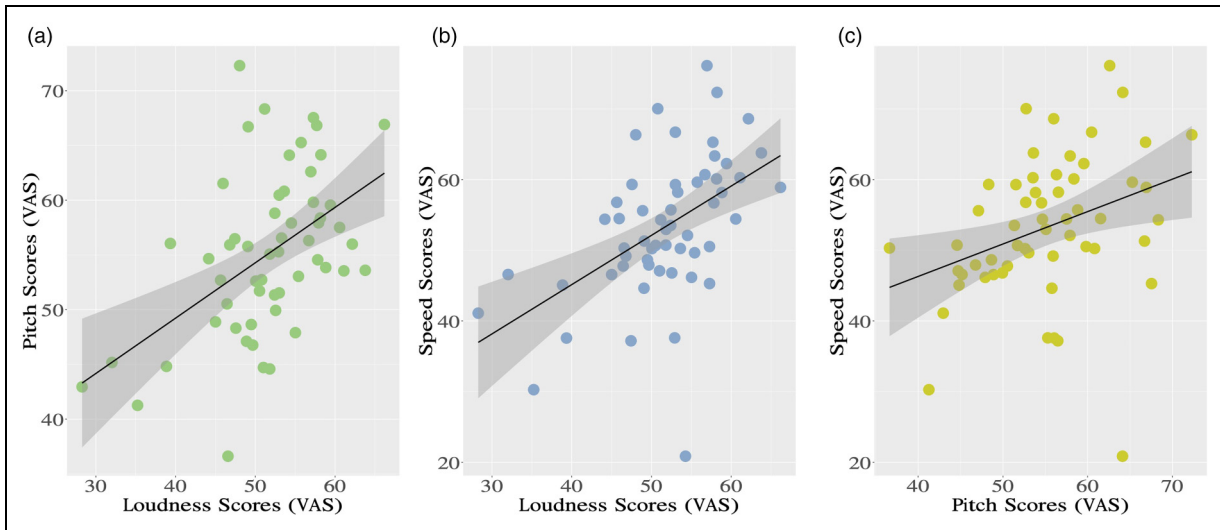
In accordance with our hypotheses, for the main effect of Vertical Density, we found that High was perceived as being the highest Pitch, followed by Low and then Medium. The same pattern of results was true for perceived Loudness in Vertical Density. A significant correlation was found between Pitch and Loudness, which may support the well-documented evidence of a common CMC between these dimensions. The present study only used a correlational design, but nevertheless this result is consistent with the previous literature, and thus it is worth considering the effects of the stimuli conditions on Pitch and Loudness together. Increased pitch height is often perceived as increased loudness (Küssner et al., 2014; Melara & Mounts, 1994), and various other correspondences have been observed between pitch height, loudness, and other modalities (Eitan, 2017), including visual elevation (e.g., Ben-Artzi & Marks, 1995), brightness (e.g., Klapetek, Ngo & Spence, 2012), physical size (e.g., Eitan et al., 2014), and tactile sensation (Peeva et al., 2004). Previous research has often used a stroop-like sensory discrimination task to measure the interference between Pitch and Loudness with increases in Pitch often being misinterpreted as increases in Loudness and vice versa (Melara & Mounts, 1994). Others have noted faster reaction times when participants are presented with stimuli with congruent attributes across dimensions (e.g., sounds that are loud and high pitch are faster to process than sounds that are loud and low pitch), according to these established CMCs (Eitan et al., 2008; Rusconi et al., 2005).

However, in the case of this study, the effect of Vertical Density on perceived Pitch and Loudness was likely observed because of the musical features specific to the Vertical Density variable in the stimuli. Low consisted of only a single melodic line, Medium added a pedal note below the melodic line, and High added additional melodic voices. It seems that participants perceived the pedal note as lowering the overall Pitch of the stimulus in the Medium condition, while the upper voices in the High condition increased the overall sense of Pitch height. Similarly for Loudness, the additional voices likely added more acoustic energy. This is confirmed with the RMS energy and LUFS measures and was likely perceived as an increase in Loudness.

RMS calculates the average signal power of an audio clip. LUFS differs from RMS in that it also factors in perceived loudness when measuring audio loudness, taking into account how humans naturally hear sound. Both of these measures corroborate the subjective ratings of Loudness for all conditions, suggesting that participants were perceiving a real difference in Loudness between the stimuli. Signal loudness is a physical concept, but it is also subjective across species and individual members of the population. Most species are attuned to hearing the frequencies that their con-specifics vocalize at, as these are their most ecologically relevant sounds. The composer of the stimuli used his professional expertise to maintain *mf* as average across each stimulus.

Within our stimuli, High had the highest mean RMS for Vertical Density, as already discussed, while Quadruple had the highest mean RMS for Note Pattern, and Flat had the highest mean RMS for Contour. Numerous other effects on subjective ratings of Pitch and Loudness were also observed. For the main effect of Note Pattern, increased horizontal density was associated with increased Loudness, with Quadruple being perceived as having higher Pitch as well as Loudness. For the main effect Contour, Flat was perceived as being Louder than either Ascending or Descending. These results for Loudness are consistent with objective measures; however, the subjective ratings of Pitch are not.

Although increased Vertical Density was associated with increased Loudness, there appeared to be a ceiling effect in



**Figure 12.** Correlations. Linear graphs depicting mean VAS ratings for significant correlations ( $N = 56$ ) between: (a) Loudness and Pitch, (b) Loudness and Speed, and (c) Pitch and Speed.

**Table 13.** Intrinsic features, perceptual effects, and CMCs of stimulus set.

	Stimulus feature	Perceptual effect	CMC
Contour	Flat has additional harmony	Flat is Louder	Flat is Faster
Vertical Density	High has additional, higher melodic line	High is Higher and Louder	High is Faster
Note Pattern	Quadruple has faster tempo at beat level	Quadruple is Faster	Quadruple is Louder and Higher

the two-way interactions between Vertical Density and both the Flat condition and Note Pattern, that is, the increased Loudness associated with High Vertical Density was diminished when combined with either Quadruple (Vertical Density \* Note Pattern) or Flat (Vertical Density \* Contour). These patterns are similarly reflected in the results for perceived Pitch. For the main effect of Note Pattern, Quadruple was perceived as the highest in Pitch, followed by Triple then Duple. All interactions for Pitch were significant. There seemed to be an additive effect of Vertical Density with Note Pattern as well, although with a ceiling effect again in High Vertical Density. This similarity with the results for Loudness is consistent with the strong correlation between Pitch and Loudness and suggests cross-modal effects of Note Pattern and Vertical Density.

For the main effect of Contour, specific effects on Pitch were observed. Ascending was perceived as having higher Pitch than Flat, which was perceived as having higher Pitch than Descending. However, Flat was perceived as having higher Pitch in Low Vertical Density, but not in either Medium or High. This suggests that Flat Contour is perceived as higher Pitched without the pedal note. Although Flat had a narrower Pitch range than either Ascending or Descending, Flat did have additional harmony that appears to have affected the perception of both Pitch and Loudness. This may have been exacerbated in the Quadruple condition due to the high rate of notes, as an

additional interaction effect was observed between Contour \* Note Pattern. In addition, the Flat condition does start at a higher Pitch than Ascending, and so participants may have perceived this as more consistently high Pitch. Between the Ascending and Descending conditions, Ascending did end on higher notes than Descending began in Quadruple and High conditions. This was done to maintain the melodic pattern. Nevertheless, just within the Duple, Low conditions, which were well matched for pitch range, Ascending was perceived as being higher pitch than Descending, implying that participants still perceived the increased effort in rising pitch. However, it is still possible that participants were attending to the last note in the melody to make their judgments of Pitch.

### Speed

As expected, for the main effect of Note Pattern, participants perceived Quadruple as being the highest in Speed, followed by Triple, then Duple. This confirms the potential confound observed in Study 1 between Note Pattern and perceived Speed, which may help to explain the results on perceived Movement, Rotation, and Physical Involvement in Study 1. The Quadruple condition not only had a Quadruple metrical structure, but also a higher horizontal note density, followed by Triple, then Duple. All three had a consistent tempo at the level of the bar or measure, but with more subdivisions within the bar. This is an important limitation to

note and will be expounded upon in the “General Discussion” section.

Indeed, a main effect of Vertical Density on perceived Speed was observed. This included a Vertical Density \* Note Pattern interaction, in which High Vertical Density conditions were perceived as higher in Speed compared with Medium and Low, but only for the Quadruple and Triple conditions. High conditions have an additional melodic line, which may be perceived as additional auditory events and thus an increase in Speed. As Medium conditions differ from Low with only the addition of a soft pedal note, this may explain why no significant differences were found between Low and Medium. However, this pattern was not reflected within the Duple conditions, in which the horizontal density was perhaps below the threshold for differences in Vertical Density to modulate perceived Speed. Another explanation may be the accent structure of Duple conditions, as Duple rhythms may be perceived as a walking rhythm (Larsson et al., 2019). Perhaps the relationship between Duple rhythms and walking make the perception of tempo more robust, due to their embodied nature, and less easily influenced by other factors. However, due to the confound between Note Pattern and horizontal note density (and thus Speed), it is not possible to infer this conclusively from the results.

In terms of main effects for Contour, Ascending was perceived to be faster than Descending, and Flat was perceived to be faster than both Ascending and Descending, possibly because of the additional harmony in the Flat condition. The effect of Contour interacted with Vertical Density, in that the Low Vertical Density and Flat condition was perceived as faster than either High Flat or Medium Flat conditions. High Vertical Density was perceived as Faster than Low or Medium Vertical Density, except in the Flat conditions. The Flat condition was perceived as faster than either Ascending or Descending, except within the High Vertical Density conditions. This indicates that both Flatness and High Vertical Density increase perceived Speed, although with diminishing returns when Flatness is combined with High Vertical Density. High Vertical Density may have been perceived as faster due to the increased density of musical events conveying increased energy. It is less clear why Flat conditions were perceived as faster, but it may be due to the repetitive nature of the Flat stimuli.

Significant correlations were observed between Speed and Pitch, and Speed and Loudness. Indeed, several studies have shown that higher Pitch is perceived as moving faster than lower Pitch (Eitan, 2013). Walker and Smith (1984, 1986) showed that differentiation of the words “fast” and “slow” is faster when the words are accompanied by high and low tones, respectively. In an adjective rating experiment, participants rated the word “fast” as more appropriate to a higher-pitched musical excerpt, and “slow” as more appropriate to a lower-pitched one (Eitan & Timmers, 2010). In dynamic pitch, however, descent is associated with acceleration, rather than

deceleration (Eitan & Granot, 2006), although the contrary was reported in the current study. Loudness and Speed may be a generalized association in listeners, acquired through experience, since increased impact velocity produces louder impact sound. In music-related imagery tasks, adults and children associated stimuli in *crescendi* with accelerating physical motion (although *diminuendi* did not evoke deceleration) in stimuli composed of equi-durational sounds (Eitan & Granot, 2006; Eitan & Tubul, 2010). The Loudness–Speed association was also demonstrated in motion tasks in which children accelerated their motion in response to *crescendi*, and vice versa (Kohn & Eitan, 2009). It must be noted that, unlike previous studies that have investigated relationships between Pitch, Loudness, and Speed, the present study included the additional dimension of Vertical Density in stimuli with multiple voices. This could explain the inconsistent results for Speed compared with previous literature.

### *Intrinsic Features vs. CMCs*

By taking into account both subjective ratings and objective measures of the stimuli, we may separate the real perceptual effects of the stimuli from illusory CMCs (Table 13). For instance, the Flat condition has additional harmony when compared with both Ascending and Descending. This may contribute to Flat being louder than the other Contour conditions on objective measures, and also in participants’ ratings of Loudness. However, participants also rated Flat as being the fastest Contour, for which there is no objective basis, suggesting that this is a CMC between Loudness and Speed. Within Vertical Density, the High condition has an additional melodic line and is objectively higher in average Pitch. This additional line also contributes to High being louder than the other Vertical Density conditions. Participants’ subjective ratings conform with the objective measures for both Pitch and Loudness, although they additionally rated High as being faster in Speed than the other conditions. Finally, Note Pattern has a confound with Speed at the beat level, as discussed earlier, due to the higher horizontal density of notes. Participants perceived Note Pattern as being associated with Speed; however, they also perceived Quadruple as being Louder and higher Pitch than Duple or Triple, which is likely a perceptual CMC.

## **General Discussion**

### *Summary of Results*

Taken together, these two studies reveal a range of embodied cross-modal associations. Study 1 tested the following theories: the EToRS, theories of general and vocal embodied responses to music, and the SAME theory. Study 2 then served to clarify the results of Study 1 and dissociate possible confounding effects in the stimuli identified in Study 1. According to the embodied music framework, musical



meaning begins with the imitation of musical sounds and the physical exertions they produce using the body (Godøy, 2003; Leman, 2007). According to Cox (2017), the voice plays a key role in embodied responses to kinetic CMCs for pitch. Meanwhile, SAME states that emotional content in music is conveyed through the Movement implied in the music, suggesting a relationship between Physical and Emotional Involvement during music listening. Finally, EToRS suggests that the perception of rotation in music should be related to metrical structure, with Triple Note Patterns being more rotatory than Duple or Quadruple (Hansen & Huron, 2019).

In terms of perceived Rotation, Study 1 found that Triple Note Patterns were perceived as rotating more than Duple patterns; however, Quadruple was higher than both. However, there was a possible confound within the stimuli used for this study, as the Note Pattern manipulation kept tempo constant at the bar level, which resulted in an increased horizontal note density within the bar or a faster tempo at the quaver level. Study 2 confirmed this, as a new set of participants rated Quadruple as being the fastest condition, followed by Triple then Duple. In addition to their findings relating to Triple Note Pattern, Hansen and Huron (2019) also found that Speed was related to Rotation. It is possible that Speed is related to Movement in general, as we observed in Study 1, which can be perceived as an increase in Rotation as well.

Results for Study 2 found a high level of interrelatedness between Pitch, Loudness, and Speed, suggesting that these may all be related to an underlying “energy” factor. This is in line with theories of embodied vocal responses to music (Cox, 2017), which link perceived pitch, effort, and perceived intensity in music. The voice is the first instrument through which most children express themselves musically, and likely the first to emerge through human evolutionary history (Bannan, 2019, p. 4). Thus, any increase in Pitch, Speed, or Loudness performed by a musical instrument (e.g., piano) may be related back to the effort that is required with the voice (Cox, 2017). Study 2 found that Ascending was perceived as being faster than Descending, despite there being no objective difference in Speed or Loudness between these conditions. In relation to Contour results for Study 1, this provides further support for theories of embodied vocal responses (Cox, 2017), as despite the low RMS values for the Ascending condition (compared with Descending and Flat), participants found Ascending to evoke more perceived Movement and more Emotional Involvement, which probably arises from the upward direction of the Contour.

Study 1 found significant correlations between Movement, Emotional Involvement, and Physical Involvement, as well as between Direction, Emotional Involvement, and Physical Involvement. The relationship between Direction and Physical Involvement is consistent with theories of vocal embodied responses mentioned earlier (Cox, 2017), as melodies perceived as moving upward may be felt as requiring more effort in the voice. Given the results of Study 2, it is

not possible to know whether participants were primarily attending to the direction of movement or the absolute pitch height of the notes when making their judgments of Physical Involvement. This could be investigated further with alterations to the stimuli.

The relationships between Movement, Emotional Involvement, and Physical Involvement provide strong support for the SAME theory, which predicts that felt emotion in music may arise through motor resonance (i.e., Physical Involvement) with the motion perceived in a musical stimulus (Overy & Molnar-Szakacs, 2009). This process could be underpinned by motor mirroring in the brain (Gallese, 2009), although no neuroimaging was conducted during this study, so the extent of motor cortex involvement is not known.

These results should be viewed within the context of general research on CMCs. Most CMC studies have not used dynamic, complex, musical stimuli (Eitan et al., 2014). By contrast, the stimuli used in this study are naturalistic including multiple voices, with the exclusion of the Low Vertical Density condition. This could explain why we found different CMCs to previous literature, particularly in relation to Speed and Pitch (Eitan & Granot, 2006). Nevertheless, most of the correspondences observed were consistent with prior research, most notably the correlation between perceived Pitch and Loudness (Melara & Mounts, 1994).

### Limitations

The results found here are consistent with an embodied account of music perception. However, there are inherent limitations in a perceptual study such as this. Without neuroimaging data to corroborate these findings, we cannot be certain of a causal relationship between motor-mimetic processes and music perception. In addition, some of our analysis was purely correlational, so the direction of causality cannot be assumed, and there is a possibility of other factors for which we have not accounted.

Designing perfectly controlled musical stimuli presents many challenges, and there were also potential confounds in the Note Pattern and Contour variables. With regard to Note Pattern, there was a confound between metrical structure and perceived Speed. Alternative stimuli could keep the tempo of subdivisions constant but change the metrical grouping (i.e., changing the accent structure within the bar), and this should be investigated in regards to perceived Rotation.

For Contour, the Flat condition relied upon the repetition of notes. This was inevitable, given that the stimuli were performed on piano. Each condition had the same number of notes, but the Flat condition had these notes repeated on the same set of pitches. There were also additional melodic lines added to the Flat condition, which appears to have created a confound with Vertical Density. This was a stylistic choice by the composer, in order to create a Flat condition that sounded “like music.” Providing a

Flat Contour with purely repeated notes would have sounded too mechanical, while playing a trill would have been perceived as fluctuating pitch, so the choice was made to have two lines moving against each other. The additional harmony in the Flat condition also allowed the Flat condition to cover a wider range of pitches to be more comparable with the pitch range of Descending and Ascending. Future studies could circumvent some of these limitations through different instrumentation. For example, a violin could use one continuous note moving in an Ascending, Descending (using *glissando*), or Flat direction. For the purposes of this study, a piano was chosen so that multiple voices from one performer could be used to investigate Vertical Density. Nevertheless, results for perceived Rotation in Study 1 indicate that Flat was perceived as being significantly less rotatory than Ascending or Descending, and previous studies (Hansen & Huron, 2019) indicate that this may be driven by a lack of moving pitch. Thus, if Flat was more repetitive, this did not appear to interfere with participants' perceptions of Flat as having an absence of Contour in Study 1.

It is also not possible to conclude whether participants felt the Direction of the Contour in terms of a vocal embodied response on a bodily level (namely, the voice), or whether they judged the Direction of the melody based on the last set of high or low notes, which are easier to discern. Indeed, participants did report Ascending to be higher than Descending. Future studies could alter the starting points of the melodies, to investigate whether some of the embodied perceptual effects are the result of pitch Contour, or absolute pitch. Although we interpret these results through the lens of theories of vocal embodied responses, based upon subjective measures of Physical Involvement, further research would also be required to determine the extent to which bodily/vocal associations are involved.

It must be noted that, while conducting an informal pilot for the Study 2 rating task questions, it became apparent that many musical terms in Italian are already embodied metaphors. Indeed, Italian music terminology denotes the expressivity of creating sounds, rather than the sounds themselves. This raises questions about the extent to which embodied metaphors in music perception may be influenced by language. In Study 1, on the other hand, some of the questions posed to participants were intentionally left vague to avoid priming effects; however, this also creates a challenge in interpreting their responses. Our findings are inherently dependent upon how the participants understood the questions, so they may not be able to contribute to a deeper theoretical understanding of motion perception in music, although Study 2 helps to clarify some of the ambiguity in the stimuli. When participants reported perceiving Movement or experiencing Physical Involvement (distinct concepts within the rating task, which nevertheless exhibited a strong relationship in analysis), we cannot be sure whether our participants understood this as metaphorical (Scruton, 1997), apparent (Gjerdingen, 1994), self-motion (Todd,

1999), or virtual (Clarke, 2001). For instance, when participants reported perceiving Rotation, we did not ask them to distinguish between whether they felt themselves rotating, or the sound source of the music. Further research would be required to separate these different understandings of perceived musical movement.

### Future Research

The present study raises additional questions that should be investigated further. Specific limitations of the stimuli have been identified, which may create opportunities for further refinement. Study 1 investigated perception of motion, and the subjective experience of Physical and Emotional Involvement. As the results here are purely from self-report measures, future studies may utilize motion capture, physiological recordings or neuroimaging to measure actual muscle activation and brain activity in response to the stimuli, and could include cross-cultural comparisons to investigate the contribution of language. There is a need for more naturalistic stimuli in perceptual and affective neuroscience studies (Saarimäki, 2021), for which the stimuli presented here could be useful candidates. The subjective ratings of Physical Involvement may be indicating an impulse to want to move to the music. It is possible that there may be a stronger mimetic response based on musical expertise, which would be consistent with neuroimaging data suggesting greater motor cortex activation when listening to familiar music (Gordon et al., 2018). However, Study 1 did not compare different levels of musical training, so this would be an avenue for further research.

In a similar vein, although there is previous literature investigating correspondences between Pitch and vertical movement (e.g., Gødoy et al., 2006), there are currently no studies on motor responses to rotating music according to the ETORS. We may expect that listening to music that affords rotation may induce an impulse for rotation, as performed in dances such as the waltz (with its  $\frac{3}{4}$  meter). One may expect that subjective ratings of Physical Involvement in response to these stimuli should also be related to motor cortex activation, as has been investigated in some existing work on motor involvement in music perception (e.g., Wallmark et al., 2017). Similarly, if vocal embodied responses contribute to simulatory CMCs between Pitch, vertical Direction, and effort (as implied through Physical Involvement and the relationships between Pitch, Speed, and Loudness observed in Study 2), then activation of the muscles around the larynx may be expected during rising melodic passages. Using neurostimulation techniques to block these relevant areas may then inhibit the experience of these associations and help to clarify the precise mechanisms behind perception of movement in music.

### Concluding Remarks

In the present study, stimuli construction and theory-testing of embodied responses to perceived musical motion are

inextricably linked. First, we test that certain cross-modal associations between Contour and vertical motion, as observed in the embodied music perception literature (Gødoy et al., 2006; Kelkar & Jensenius, 2019; Krantz et al., 2006; Küssner & Leech-Wilkinson, 2014; Küssner et al., 2014; Nymoen et al., 2011), may be partially explained by theories of embodied vocal responses (Cox, 2017). Second, that the perception of musical events as kinetic ones, tested through Vertical Density, can be explained by theories of general embodied responses (Gødoy, 2003; Leman, 2007). Third, that Note Pattern may contribute to perceptions of Rotation (Hansen and Huron, 2019). Lastly, that felt responses of Physical and Emotional Involvement would be associated (Molnar-Szakacs & Overy, 2006).

Our findings were consistent with theories of general and vocal embodied responses to music, as well as with theories of embodied emotional contagion in music. In particular, although Ascending was perceived as being evoking less Loudness than Descending and Flat according to both subject ratings in Study 2 and RMS/LUFS analysis, Study 1 subjects rated Ascending as evoking greater Movement, Physical Involvement, and Emotional Involvement, suggesting that Study 1 results for perceived motion in Contour were in response to moving pitch. The reported results suggest that many cross-modal associations may originate from feelings of effort, supported in part by a correlational relationship between Pitch, Loudness, and Speed in Study 2 and the relationship between Movement, Direction, and Physical Involvement in Study 1. Therefore, any musical factor that affords greater effort, whether through increased harmonic complexity, tempo, or rising pitch, may increase feelings of physical exertion. The CMC of increased intensity associated with rising pitch (“greater is higher”) may then be felt as increased emotional involvement.

However, the follow-up investigation (Study 2) identified confounds due to the intrinsic features of the stimuli and CMCs within the stimuli, in particular within the Contour, Note Pattern, and Vertical Density variables. In particular, further studies balancing accent patterns with note density are needed to investigate perceived musical Rotation further. In creating controlled stimuli, certain compromises are needed and future research could build upon the stimulus set to address these factors. Despite its limitations, the stimulus set is well positioned to serve researchers seeking to balance experimental control for numerous variables while maintaining aesthetic qualities that are still “musical” (available through open access). Finally, further research using neuroimaging, neurostimulation or physiological measures is required to investigate further and confirm the reported results.

### Acknowledgements

The authors would like to thank Gioacchino Garofalo and Francesca Siri for their help with the research design and creation of the Python script, Annalisa Pelosi for her help with the data analysis, and the Reviewers for their insightful and valuable feedback.

### Action Editor

Ian Cross, University of Cambridge, Faculty of Music

### Peer Review

Two anonymous reviewers

### Contributorship

Conceived of the experiment: AK, JB, EA; designed the experiment: AK, JB, EA, MM, MC, MAU; performed the experiment: AK, MM, MP; analyzed the data: AK, NL; wrote the initial draft: AK and MAU; initiated the project: AK, MC, MAU, VG, MG. All authors contributed substantially to the revision of the initial draft and approval of the final version of the manuscript.

### Declaration of Conflicting Interests

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

### Funding


The authors disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: Research Council of Finland (346210 and 332331), Cariparma Foundation.

### Ethical Approval

The Area Vasta Emilia Nord (AVEN) Ethics Committee approved this study (REF: 85/2019/DISP/UNIPR).

### ORCID iDs

Anna Kolesnikov  <https://orcid.org/0000-0002-3650-9306>

Joshua S. Bamford  <https://orcid.org/0000-0002-8204-7915>

### Supplemental Material

Supplemental material for this article is available online.

### References

- Albiero, P., Ingoglia, S., & Lo Coco, A. (2006). Contributo all'adattamento Italiano dell' Interpersonal Reactivity Index TT - A contribution to the Italian validation of the Interpersonal Reactivity Index. *Test. Psicometria Metodol.*
- Bamford, J. S., & Davidson, J. W. (2019). Trait empathy associated with agreeableness and rhythmic entrainment in a spontaneous movement to music task: Preliminary exploratory investigations. *Musicae Scientiae*, 23(1), 5–24. <https://doi.org/10.1177/1029864917701536>
- Bannan, N. (2019). *First Instruments: Teaching Music Through Harmony Singing*. Oxford University Press.
- Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, 67(1), 1–48. <https://doi.org/10.18637/jss.v067.i01>
- Ben-Artzi, E., & Marks, L. E. (1995). Visual-auditory interaction in speeded classification: Role of stimulus difference. *Perception & Psychophysics*, 57, 1151–1162. <https://doi.org/10.3758/BF03208371>

- Brodsky, W., Kessler, Y., Rubinstein, B.-S., Ginsborg, J., & Henik, A. (2008). The mental representation of music notation: Notational audition. *Journal of Experimental Psychology: Human Perception and Performance*, *34*(2), 427–445. <https://doi.org/10.1037/0096-1523.34.2.427>
- Burger, B., Thompson, M. R., Luck, G., Saarikallio, S., & Toiviainen, P. (2012). Music moves US: beat-related musical features influence regularity of music-induced movement. In E. Cambouropoulos, C. Tsougras, P. Mavromatis, & K. Pasiadis (Eds.), *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*, Thessaloniki, Greece.
- Callan, D. E., Tsytsarev, V., Hanakawa, T., Callan, A. M., Katsuhara, M., Fukuyama, H., & Turner, R. (2006). Song and speech: Brain regions involved with perception and covert production. *Neuroimage*, Elsevier, *31*(3), 1327–1342. <https://doi.org/10.1016/j.neuroimage.2006.01.036>
- Chen, J. L., Penhune, V. B., & Zatorre, R. J. (2008a). Listening to musical rhythms recruits motor regions of the brain. *Cerebral cortex*, Oxford University Press, *18*(12), 2844–2854. <https://doi.org/10.1093/cercor/bhn042>
- Chen, J. L., Penhune, V. B., & Zatorre, R. J. (2008b). Moving on time: Brain network for auditory-motor synchronisation is modulated by rhythm complexity and musical training. *Journal of Cognitive Neuroscience*, MIT Press, *20*(2), 226–239. <https://doi.org/10.1162/jocn.2008.20018>
- Christensen, R. H. B. (2019). A Tutorial on fitting Cumulative Link Mixed Models with `clmm2` from the ordinal Package. *Tutorial for the R Package ordinal*. <https://cran.r-project.org/web/packages/ordinal/index.html>
- Clarke, E. (2001). Meaning and the specification of motion in music. *Musicae Scientiae*, *5*(2), 213–234. <https://doi.org/10.1177/102986490100500205>
- Clemente, A., Vila-Vidal, M., Pearce, M. T., Aguiló, G., Corradi, G., & Nadal, M. (2020). A set of 200 musical stimuli varying in balance, contour, symmetry, and complexity: Behavioural and computational assessments. *Behaviour Research Methods*, *52*(2020), 1491–1509. <https://doi.org/10.3758/s13428-019-01329-8>
- Core, T. R. (2019). *R: A language and environment for statistical computing*. Vienna, Austria.
- Cox, A. (2001). The mimetic hypothesis and embodied musical meaning. *Musicae scientiae*, SAGE Publications Sage UK: London, England, *5*(2), 195–212. <https://doi.org/10.1177/102986490100500204>
- Cox, A. (2011). Embodying music: Principles of the mimetic hypothesis. *Music Theory Online*, *17*(2).
- Cox, A. (2017). *Music and embodied cognition: Listening, moving, feeling, and thinking. Illustrated edition*. Indiana University Press.
- Davis, M. H. (1980). A multidimensional approach to individual differences in empathy. American Psychological Association Washington, DC.
- Del Negro, C. A., Funk, G. D., & Feldman, J. L. (2018). Breathing matters. *Nature Reviews Neuroscience*, *19*(2018), 351–367. <https://doi.org/10.1038/s41583-018-0003-6>
- Di Pellegrino, G., Fadiga, L., Fogassi, L., Gallese, V., & Rizzolatti, G. (1992). Understanding motor events: A neurophysiological study. *Experimental Brain Research*, *91*, 176–180. <https://doi.org/10.1007/BF00230027>
- 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>
- Eerola, T., & Vuoskoski, J. K. (2011). A comparison of the discrete and dimensional models of emotion in music. *Psychology of Music*, *39*(1), 18–49. <https://doi.org/10.1177/0305735610362821>
- Eitan, Z. (1997). *Highpoints: a study of melodic peaks*. Philadelphia, Pennsylvania. University of Pennsylvania Press.
- Eitan, Z. (2007). Intensity and cross-dimensional interaction in music: Recent research and its implications for performance studies. *Orbis Musicae: Studies in Musicology*, (14), 141–166.
- Eitan, Z. (2013). How pitch and loudness shape musical space and motion. In S.-L. Tan, A. J. Cohen, & S. D. Lipscomb, & R. A. Kendall (Eds.), *The psychology of music in multimedia* (pp. 165–191). Oxford University Press. <https://doi.org/10.1093/acprof:oso/9780199608157.003.0008>
- Eitan, Z. (2017). Crossmodal correspondences. In R. Ashley & R. Timmers (Eds.), *The Routledge companion to music cognition* (pp. 214–221). Routledge.
- Eitan, Z., & Granot, R. Y. (2006). How music moves: Musical parameters and listeners' images of motion. *Music Perception*, *23*(3), 221–247. <https://doi.org/10.1525/mp.2006.23.3.221>
- Eitan, Z., Schupak, A., Gotler, A., & Marks, L. E. (2014). Lower pitch is larger, yet falling pitches shrink: Interaction of pitch change and size change in speeded discrimination. *Experimental Psychology*, *61*(4), 273–284. <https://doi.org/10.1027/1618-3169/a000246>
- Eitan, Z., Schupak, A., & Marks, L. E. (2008). Louder is higher: cross-modal interaction of loudness change and vertical motion in speeded classification.
- 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. <https://doi.org/10.1016/j.cognition.2009.10.013>
- Eitan, Z., & Tubul, N. (2010). Musical parameters and children's images of motion. *Musicae Scientiae, Special Issue, 2010*, 89–111. <https://doi.org/10.1177/10298649100140S207>
- Faul, F., Erdfelder, E., Lang, A. G., & Buchner, A. (2007). G\* Power 3: A flexible statistical power analysis program for the social, behavioural, and biomedical sciences. *Behaviour Research Methods*. Springer, *39*(2), 175–191. <https://doi.org/10.3758/BF03193146>
- Fox, J. (2003). Effect displays in R for generalised linear models. *Journal of Statistical Software*, *8*(15), 1–27. <https://doi.org/10.18637/jss.v008.i15>
- Gabrielsson, A. (2002). Emotion perceived and emotion felt: Same or different? *Musicae Scientiae*, *5*, 123–127. <https://doi.org/10.1177/10298649020050S105>
- Gallese, V. (2005). Embodied simulation: From neurons to phenomenal experience. *Phenomenology and the Cognitive Sciences*, *4*, 23–48. <https://doi.org/10.1007/s11097-005-4737-z>

- Gallese, V. (2009). Motor abstraction: A neuroscientific account of how action goals and intentions are mapped and understood. *Psychological Research PRPF*, 73(4), 486–498. <https://doi.org/10.1007/s00426-009-0232-4s>
- Gallese, V. (2014). Bodily selves in relation: Embodied simulation as second-person perspective on intersubjectivity. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 369(1644), 20130177. <https://doi.org/10.1098/rstb.2013.0177>
- Gallese, V., Fadiga, L., Fogassi, L., & Rizzolatti, G. (1996). Action recognition in the premotor Cortex. *Brain*, 119(2), 593–609. <https://doi.org/10.1093/brain/119.2.593>
- Gallese, V., & Sinigaglia, C. (2011). What is so special about embodied simulation? *Trends in Cognitive Sciences*, 15(11), 512–519. <https://doi.org/10.1016/j.tics.2011.09.003>
- Gibson, J. J. (1966). *The senses considered as perceptual systems*. Houghton Mifflin.
- Gibson, J. J. (1979). *The ecological approach to visual perception*. Houghton, Mifflin and Company.
- Gjerdingen, R. O. (1994). Apparent motion in music? *Music Perception*, 11(4), 335–370. <https://doi.org/10.2307/40285631>
- Gødøy, R. I. (2003). Motor-mimetic music cognition. *Leonardo*, 36(4), 317–319. <https://doi.org/10.1162/002409403322258781>
- Gødøy, R. I. (2008). Reflections on chunking in music. In A. Schneider (Ed.), *Systematic and comparative musicology: Concepts, methods, findings* (pp. 117–131). Peter Lang.
- Gødøy, R. I., Haga, E., & Jensenius, A. R. (2006). Exploring music-related gestures by sound-tracing. A preliminary study.
- Gødøy, R. I., & Leman, M. (2010). *Musical gestures: Sound, movement, and meaning* (1st edition). Routledge.
- Gordon, C. L., Cobb, P. R., & Balasubramaniam, R. (2018). Recruitment of the motor system during music listening: An ALE meta-analysis of fMRI data. *PloS one*, 13(11), e0207213. <https://doi.org/10.1371/journal.pone.0207213>
- Grahn, J. A., & Brett, M. (2007). Rhythm and beat perception in motor areas of the brain. *Journal of Cognitive Neuroscience*, 19(5), 893–906. <https://doi.org/10.1162/jocn.2007.19.5.893>
- Halpern, A. R., & Zatorre, R. J. (1999). When that tune runs through your head: A PET investigation of auditory imagery for familiar melodies. *Cerebral cortex*, Oxford University Press, 9(7), 697–704. <https://doi.org/10.1093/cercor/9.7.697>
- Halpern, A. R., Zatorre, R. J., Bouffard, M., & Johnson, J. A. (2004). Behavioural and neural correlates of perceived and imagined musical timbre. *Neuropsychologia. Elsevier*, 42(9), 1281–1292. <https://doi.org/10.1016/j.neuropsychologia.2003.12.017>
- Hansen, N. C., & Huron, D. (2019). Twirling triplets: The qualia of rotation and musical rhythm. *Music & Science*. SAGE Publications Sage UK: London, England, 1–20. <https://doi.org/10.1177/2059204318812243>
- Hanslick, E. (1891). *The beautiful in music: a contribution to the revisal of musical aesthetics*. Novello.
- Hickok, G., Buchsbaum, B., Humphries, C., & Muftuler, T. (2003). Auditory–motor interaction revealed by fMRI: Speech, music, and working memory in area Spt. *Journal of Cognitive Neuroscience*, MIT Press, 15(5), 673–682. <https://doi.org/10.1162/089892903322307393>
- Hodges, D. A. (2009). Bodily responses to music. In S. Hallam, I. Cross, & M. Thaut (Eds.), *The Oxford handbook of music psychology* (pp. 121–130). Oxford University Press.
- Iacoboni, M. (2009). Imitation, empathy, and mirror neurons. *Annual Review of Psychology*, 60(1), 653–670. <https://doi.org/10.1146/annurev.psych.60.110707.163604>
- Johnson, M. (2007). *The meaning of the body: Aesthetics of human understanding*. University of Chicago Press.
- Juslin, P. N., & Laukka, P. (2004). Expression, perception and induction of musical emotions: A review and a questionnaire study of everyday listening. *Journal of New Music Research*, 33(3), 217–238. <https://doi.org/10.1080/0929821042000317813>
- Juslin, P. N., & Västfjäll, D. (2008). Emotional responses to music: The need to consider underlying mechanisms. *Behavioural and Brain Sciences*, 31, 559–621. <https://doi.org/10.1017/S0140525X08005293>
- Kelkar, T., & Jensenius, A. R. (2019). Exploring melody and motion features in “sound-tracings”. Proceedings of the 14th Sound and Music Computing Conference 2017, SMC 2017, 98–103.
- 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, 1154–1167. <https://doi.org/10.3758/s13414-012-0317-9>
- Kohn, D., & Eitan, Z. (2009). Musical parameters and children’s movement responses. Proceedings of the 7th Triennial Conference of European Society for the Cognitive Sciences of Music (ESCOM 2009), Jyväskylä, Finland.
- Krantz, G., Madison, G., & Merker, B. (2006). Melodic intervals as reflected in body movement. <9th International Conference on Music Perception and Cognition, University of Bologna.
- Küssner, M. B., & Leech-Wilkinson, D. (2014). Investigating the influence of musical training on cross-modal correspondences and sensorimotor skills in a real-time drawing paradigm. *Psychology of Music*, 42(3), 448–469. <https://doi.org/10.1177/0305735613482022>
- Küssner, M. B., Tidhar, D., Prior, H. M., & Leech-Wilkinson, D. (2014). Musicians are more consistent: Gestural cross-modal mappings of pitch, loudness and tempo in real-time. *Frontiers in Psychology*, 5, 789. <https://doi.org/10.3389/fpsyg.2014.00789>
- Laban, R. (1947). *Effort*. MacDonald & Evans.
- Lakoff, G., & Johnson, M. (2008). *Metaphors We Live By*. University of Chicago Press.
- Larsson, M., Richter, J., & Ravignani, A. (2019). Bipedal Steps in the Development of Rhythmic Behavior in Humans. *Music & Science*, 2, <https://doi.org/10.1177/2059204319892617>
- Lartillot, O., & Toiviainen, P. (2007). A Matlab Toolbox for Musical Feature Extraction From Audio. International Conference on Digital Audio Effects, Bordeaux.
- Leman, M. (2007). *Embodied music: cognition and mediation technology*. MIT Press.
- Lenth, R. V., Buerkner, P., Herve, M., Love, J., Riebl, H., & Singmann, H. (2020). *emmeans: Estimated marginal means, aka least-squares means, v1. 5.1*. R Core Team.
- Maes, P. J., van Dyck, E., Lesaffre, M., Leman, M., & Kroonenberg, P. M. (2014). The coupling of action and perception in musical meaning formation. *Music Perception*, 32(1), 67–84. <https://doi.org/10.1525/mp.2014.32.1.67>

- Melara, R. D., & Mounts, J. R. W. (1994). Contextual influences on interactive processing: Effects of discriminability, quantity, and uncertainty. *Perception & Psychophysics*, *56*, 73–90. <https://doi.org/10.3758/BF03211692>
- Molnar-Szakacs, I., & Overy, K. (2006). Music and mirror neurons: From motion to ‘e’motion. *Social Cognitive and Affective Neuroscience*, *1*(3), 235–241. <https://doi.org/10.1093/scan/nsl029>
- Mondloch, C. J., & Maurer, D. (2004). Do small white balls squeak? Pitch-object correspondences in young children. *Cognitive, Affective & Behavioural Neuroscience*, *4*(2), 133–136. <https://doi.org/10.3758/CABN.4.2.133>
- Moorthigari, V., Carlson, E., Toiviainen, P., Vuust, P., Brattico, E., & Alluri, V. (2021). Trait empathy modulates music-related functional connectivity. *BioRxiv*. <https://doi.org/10.1101/2021.08.18.456484>
- Ndebele, P. (2013). The declaration of Helsinki, 50 years later. *Jama*, American Medical Association, *310*(20), 2145–2146. <https://doi.org/10.1001/jama.2013.281316>
- Nistri, A., Ostroumov, K., Sharifullina, E., & Taccola, G. (2006). Tuning and playing a motor rhythm: How metabolic glutamate receptors orchestrate generation of motor patterns in the mammalian central nervous system. *The Journal of Physiology*, *572*(2), 323–344. <https://doi.org/10.1113/jphysiol.2005.100610>
- Nymo, K., Caramiaux, B., Kozak, M., & Torresen, J. (2011). Analysing sound tracings - a multimodal approach to music information retrieval. *MIRUM'11*. Scottsdale, Arizona, USA.
- Oldfield, R. C. (1971). The assessment and analysis of handedness: The Edinburgh inventory. *Neuropsychologia*. *Oxford*, *9*(1), 97–113. [https://doi.org/10.1016/0028-3932\(71\)90067-4](https://doi.org/10.1016/0028-3932(71)90067-4)
- Overy, K., & Molnar-Szakacs, I. (2009). Being together in time: Musical experience and the mirror neuron system. *Music Perception*, *26*(5), 489–504. <https://doi.org/10.1525/mp.2009.26.5.489>
- Parise, C. V., Spence, C., & Deroy, O. (2016). Understanding the correspondences: Introduction to the special issue on cross-modal correspondences. *Multisensory Research*, *29*(1-3), 1–6. <https://doi.org/10.1163/22134808-00002517>
- Peeva, D., Baird, B., Izmirli, O., & Blevins, D. (2004). Haptic and sound correlations: Pitch, loudness and texture. Proceedings. Eighth International Conference on Information Visualisation, London, UK, 659–664. <https://doi.org/10.1109/IV.2004.1320212>
- Pearce, J., Gray, J., Simpson, S., MacAskill, M., Höchenberger, R., Sogo, H., Kastman, E., & Lindeløv, J. (2019). PsychoPy2: Experiments in behaviour made easy. *Behaviour Research Methods*. <https://doi.org/10.3758/s13428-018-01193-y> 195–203.
- Pettersen, V., Bjørkøy, K., Torp, H., & Westgaard, R. H. (2005). Neck and shoulder muscle activity and thorax movement in singing and speaking tasks with variation in vocal loudness and pitch. *Journal of Voice*, *19*(4), 623–634. <https://doi.org/10.1016/j.jvoice.2004.08.007>
- Pratt, C. C. (1931). *The meaning of music - A study in psychological aesthetics*. Johnson Reprint Corporation.
- Repp, B. (1992). Music as Motion: A Synopsis of Alexander Truslit's (1938). *Haskins Laboratories Status Report on Speech Research*, 265–278.
- (2010). forte. In Bulhosen, P., Logi, F., & Riu, L. (Eds.), *Pocket Oxford Italian Dictionary: Italian-English*. Oxford University Press. Retrieved 6 Feb. 2023, from <https://www.oxfordreference.com/view/10.1093/acref/9780191739569.001.0001/b-it-en-00002-0010064>
- Rizzolatti, G., Fadiga, L., Gallese, V., & Fogassi, L. (1996). Premotor cortex and the recognition of motor actions. *Cognitive Brain Research*, 131–141. [https://doi.org/10.1016/0926-6410\(95\)00038-0](https://doi.org/10.1016/0926-6410(95)00038-0)
- Roberts, R., Callow, N., Hardy, L., Markland, D., & Bringer, J. (2008). Movement imagery ability: Development and assessment of a revised version of the vividness of movement imagery questionnaire. *Journal of Sport and Exercise Psychology*, Human Kinetics, Inc., *30*(2), 200–221. <https://doi.org/10.1123/jsep.30.2.200>
- Rusconi, E., Kwan, B., Giordano, B. L., Umiltà, C., & Butterworth, B. (2005). Spatial representation of pitch height: The SMARC effect. *Cognition*, *99*, 113–129. <https://doi.org/10.1016/j.cognition.2005.01.004>
- Saarikallio, S., Luck, G., Burger, B., Thompson, M., & Toiviainen, P. (2013). Dance moves reflect current affective state illustrative of approach-avoidance motivation. *Psychology of Aesthetics, Creativity and the Arts*, *7*(3), 296–305. <https://doi.org/10.1037/a0032589>
- Saarimäki, H. (2021). Naturalistic stimuli in affective neuroimaging: a review. *Frontiers in Human Neuroscience*, *15*, <https://doi.org/10.3389/fnhum.2021.675068>
- Schenker, H. (1954). *Harmony*. University of Chicago Press.
- Scruton, R. (1997). *The aesthetics of music*. Oxford University Press.
- Sessions, R. (1941). *The composer and his message In: The intent of the artist*. Edited by A. Centeno. Princeton University Press.
- Shafir, T., Tsachor, R. P., & Welch, K. B. (2016). Emotion Regulation through Movement: Unique Sets of Movement Characteristics are Associated with and Enhance Basic Emotions. *Frontiers in Psychology*, *6*, <https://doi.org/10.3389/fpsyg.2015.02030>
- Sievers, B., Polansky, L., Casey, M., & Wheatley, T. (2013). Music and movement share a dynamic structure that supports universal expressions of emotion. *Proceedings of the National Academy of Sciences of the United States of America*, *110*(1), 70–75. <https://doi.org/10.1073/pnas.1209023110>
- Spence, C., & Deroy, O. (2012). Crossmodal correspondences: Innate or learned? *i-Perception*, *3*, 316–318. <https://doi.org/10.1068/i0526ic>
- Stern, D. N. (2010). *Forms of vitality: Exploring dynamic experience in psychology and the arts. Illustrated edition*. Oxford University Press.
- Stewart, L., Walsh, V., & Frith, U. (2004). Reading music modifies spatial mapping in pianists. *Perception & Psychophysics*, *66*(2), 183–195. <https://doi.org/10.3758/BF03194871>
- Timmers, R. (2017). Emotion in Music Listening. In R. Ashley & R. Timmers (Eds.), *The Routledge companion to music cognition*. Routledge.
- Todd, N. P. M. (1999). Motion in music: A neurobiological perspective. *Music Perception*, *17*(1), 115–126. <https://doi.org/10.2307/40285814>
- Tyack, P. L. (2016). Vocal learning and auditory-vocal feedback. In *Vertebrate sound communication* (pp. 261–295). Springer.

- Walker, P., Bremner, J. G., Mason, U., Spring, J., Mattock, K., Slater, A., & Johnson, S. P. (2010). Preverbal infants' sensitivity to synaesthetic cross-modality correspondences. *Psychological Science, 21*(1), 21–25. <https://doi.org/10.1177/0956797609354734>
- Walker, P., & Smith, S. (1984). Stroop interference based on the synaesthetic qualities of auditory pitch. *Perception, 13*, 75–81. <https://doi.org/10.1068/p130075>
- Walker, P., & Smith, S. (1986). The basis of stroop interference involving the multimodal correlates of auditory pitch. *Perception, 15*, 491–496. <https://doi.org/10.1068/p150491>
- Wallmark, Z., Deblieck, C., & Iacoboni, M. (2018). Neurophysiological effects of trait empathy in music listening. *Frontiers in Behavioural Neuroscience*. <https://doi.org/10.3389/fnbeh.2018.00066>
- Wallmark, Z., Iacoboni, M., Deblieck, C., & Kendall, R. A. (2017). Embodied listening and timbre: Perceptual, acoustical, and neural correlates. *Music Perception, 25*(3), 332–363. <https://doi.org/10.1525/mp.2018.35.3.332>
- Watt, R. J., & Ash, R. L. (1998). A psychological investigation of meaning in music. *Musicae Scientiae, 2*(1), 33–53. <https://doi.org/10.1177/102986499800200103>
- Werner, P. D., Swope, A. J., & Heide, F. J. (2006). The music experience questionnaire: Development and correlates. *The Journal of Psychology. Taylor & Francis, 140*(4), 329–345. <https://doi.org/10.3200/JRLP.140.4.329-345>
- Wickham, H. (2016). *ggplot2: Elegant graphics for data analysis*. Springer.
- Zatorre, R. J., Halpern, A. R., Perry, D. W., Meyer, E., & Evans, A. C. (1996). Hearing in the mind's ear: A PET investigation of musical imagery and perception. *Journal of Cognitive Neuroscience. MIT Press, 8*(1), 29–46. <https://doi.org/10.1162/jocn.1996.8.1.29>
- Zuckermandl, V. (1956). *Sound and symbol: Music and the external world*. Pantheon.
- Žuvela, S. K., & Anić, A. O. (2019). The embodied and the cultural in the conceptualization of pitch space in Croatian. *Jezikoslovlje, 20*(2), 199–219. <https://doi.org/10.29162/jez.2019.7>