Beethoven’s last piano sonata and those who follow crocodiles: Cross-domain mappings of auditory pitch in a musical context

Zohar Eitan a,*, Renee Timmers b

a School of Music, Tel Aviv University, 69978 Tel Aviv, Israel
b Department of Music, University of Sheffield, United Kingdom

Abstract

Though auditory pitch is customarily mapped in Western cultures onto spatial verticality (high–low), both anthropological reports and cognitive studies suggest that pitch may be mapped onto a wide variety of other domains. We collected a total number of 35 pitch mappings and investigated in four experiments how these mappings are used and structured. In particular, we inquired (1) how Western subjects apply Western and non-Western metaphors to “high” and “low” pitches, (2) whether mappings applied in an abstract conceptual task are similarly applied by listeners to actual music, (3) how mappings of spatial height relate to these pitch mappings, and (4) how mappings of “high” and “low” pitch associate with other dimensions, in particular quantity, size, intensity and valence. The results show strong agreement among Western participants in applying familiar and unfamiliar metaphors for pitch, in both an abstract, conceptual task (Exp. 1) and in a music listening task (Exp. 2), indicating that diverse cross-domain mappings for pitch exist latent besides the common verticality metaphor. Furthermore, limited overlap between mappings of spatial height and pitch height was found, suggesting that, the ubiquity of the verticality metaphor in Western usage notwithstanding, cross-domain pitch mappings are largely independent of that metaphor, and seem to be based upon other underlying dimensions. Part of the discrepancy between spatial height and pitch height is that, for pitch, “up” is not necessarily “more,” nor is it necessarily “good.” High pitch is only “more” for height, intensity and brightness. It is “less” for mass, size and quantity. We discuss implications of these findings for music and speech prosody, and their relevance to notions of embodied cognition and of cross-domain magnitude representation.

Keywords:
Cross-domain mapping
Metaphor
Auditory pitch
Music cognition

1. Introduction

Discourse concerning auditory phenomena, and music specifically, relies heavily on terms derived from non-auditory realms of experience, applying metaphorical mappings from visuo-spatial, kinaesthetic, or tactile domains. We habitually speak, for instance, of high and low (or ascending and descending) pitches, bright and dark timbres of an instrument or voice, rough sounds, and heavy or light-footed rhythms.

A large number of empirical studies have shown that such cross-domain mappings are not merely convenient figures of speech, but influence auditory perception, as demonstrated in experiments involving selective attention, object discrimination, and similarity perception (see summaries in Marks, 1996, 2000, 2004). Indeed, cross-modal associations, including, for instance, interactions between loudness and brightness, were reported even for three months old infants, suggesting that non-linguistic, probably innate perceptual equivalences may be involved in the formation of some metaphoric mappings of sound (Lewkowicz & Turkewitz, 1980).

For auditory pitch, the central cross-domain association in Western culture maps “high” and “low” pitch onto the
corresponding spatial poles (the “verticality” metaphor). This analogy is not only pervasive in language, but affects auditory and spatial perception, as well as speech production (Shintel, Nusbaum, & Okrent, 2006). Higher pitch is actually perceived as emitted from a higher location in space (Bregman & Steiger, 1980; Cabrera, Ferguson, Tilley, & Morimoto, 2005; Pratt, 1930; Roffler & Butler, 1968; Timble, 1934); correspondingly, the height of concurrent vertical stimuli affects the estimation of pitch “height” (Casasanto, Phillips, & Boroditsky, 2003). Congruence of pitch height with vertical position of visual stimuli also affects response time in auditory and visual speeded discrimination tasks (Ben-Artzi & Marks, 1995; Bernstein & Edelstein, 1971; Melara & O’Brien, 1987) as well as in stimulus–response compatibility experiments (Lidji, Kolinsky, Lochy, & Morais, 2007; Rusconi, Kwan, Giordano, Umita, & Butterworth, 2005; Simon, Mewaldt, Acosta, & Hu, 1976). The direction of pitch change has even been shown to affect infants’ attention to corresponding visual stimuli, suggesting that this cross-modal analogy might be inborn (Wagner, Winner, Cicchetti, & Gardner, 1981; Walker et al., in press).

The pervasiveness of the pitch-verticality analogy notwithstanding, it is neither unique nor universal. As evidence based on several experimental paradigms indicates, mappings of pitch to domains other than verticality (not all of which are expressed in verbal discourse) partake in its perception and cognition. In experiments utilizing the Stroop paradigm, matching high or low pitches with verbal antonyms, higher pitch was found to be congruent with the adjectives small, bright, sharp, fast, and active, and lower pitch – with their opposites (Walker & Smith, 1984, p. 86). In speeded discrimination experiments presenting high and low pitches simultaneously with visual stimuli, higher pitch concurred with brighter light, lighter color, and sharper edges (Marks, 1982, 1987). Pitch was also related to spatio-kinetic dimensions other than rise and fall. In music imagery experiments, pitch “rise” has been associated with spatial rise, but also with distancing, acceleration, and increasing energy; pitch “fall,” while associated with spatial fall, was also related to approaching, turning left, and decreasing energy (Eitan & Granot, 2006).

Besides influences of pitch associations on perception, cross-domain mappings of pitch also play a role in speech production. Together with other acoustic cues, such as duration, loudness and pitch variation, pitch height (F0) was found to be an important semantic marker in speech. Nygaard, Herold, and Namya (2009) asked participants to produce phrases in infant-directed speech in which novel words were assigned one of two meanings in several antonym pairs. Speakers assigned specific acoustic features to different word meanings, and these features were used by other participants to reliably identify the intended meanings of the novel words. Pitch height was used in the antonyms happy–sad, yummy–Yucky, big–small, and hot–cold, where higher pitch was associated with happy, yummy, small and cold. More generally, higher pitch was often correlated with positive evaluation.

Cross-cultural anthropological studies also indicate a diversity of connotations for pitch. In different cultures and historical eras, pitch polarity was not designated as “high” vs. “low” but rather by “light” vs. “heavy” (Kpelle people in Liberia; see Stone, 1981). “Sharp” vs. “heavy” (ancient Greek music theory), “small” vs. “large,” used in Bali and Java, as well as among Kpelle and Jabo in Liberia (Stone, 1981), “young” vs. “old” (Suyá people of the Amazon basin; see Zbikowski, 1998) or “weak” vs. “strong” (the Bashi people of central Africa; Merriam, 1964). Often, pitch vocabulary seems to derive from specific cultural practices. For instance, pitch classification for the Shona mbira (Zimbabwe) includes the opposition of “crocodile” (low pitch) with “those who follow crocodiles” (high), and “stable (person) who holds the piece together” (low) vs. “mad person” (high), as well as “old men’s voices” (low) vs. “young men’s voices” (high), “men’s voices” vs. “women’s voices,” and “thin” (low) vs. “thick” (high) (Berliner, 1978, discussed in Ashley, 2004). In the Gbaya xylophone (Central African Republic), notes are arranged genealogically, and include (from low to high) grandmother, mother, father, son and daughter (Ashley, 2004; after Arom & Voisin, 1998).

Such cross-cultural diversity seems to suggest that pitch mappings are learned and dependent on acculturation. However, this does not necessarily mean that the structure underlying diverse pitch metaphors shows equal diversity. Instead, pitch metaphors may be retracted to a limited set of underlying mappings, perhaps stemming from basic bodily experiences and interactions with the physical environment, or even, as Smith and Walker (1986) suggest, on a single abstract semantic code, that comprises specific oppositions in diverse dimensions. This paper aims at increasing insight into the multidimensionality of pitch metaphors and their possible cross-cultural underpinnings.

One possible conceptual basis for pitch mappings is the verticality metaphor itself, as the association of pitch with verticality may align it indirectly with other domains, which themselves are mapped metaphorically onto spatial high and low. As Lakoff and Johnson (1980a, 1980b), following Nagy (1974) show, the up–down relationship serves as a core for a large-scale “orientational” metaphoric system, encompassing conceptual metaphors such as “up is more, down is less” “happy is up, sad is down,” or “good is up, bad is down.” Importantly, such orientational metaphors have been shown to affect not only verbal, but also non-verbal and non-linguistic behavior. Thus, utterances of positive or negative valence, even when not including verticality metaphors (e.g., “my grades got better”) were accompanied, without the performer’s awareness, by congruent up and down motions (Casasanto, 2008). Correspondingly, moving objects up or down enhanced the retrieval of positively and negatively valenced memories, respectively (Casasanto & Dijkstra, submitted for publication). Automatic mapping of valenced terms into spatial verticality was also shown by priming effects in the other direction: presenting negative or positive words as primes facilitated discrimination between targets, such as the letters p and q, which were presented in lower or higher spatial positions, respectively (Meier & Robinson, 2004; see also Meier and Robinson (2005), for a survey of other relevant empirical research).
Such implicit and automatic correspondence of positions and directions on the vertical spatial plane with non-verbal behavior may engender a “second-order” mapping of “high” and “low” auditory pitch into features such as valence, mood or social hierarchy, as well as physical features like size and mass. Low pitch, for instance, may be indirectly associated with negative evaluation through its mapping into spatial height, as the latter commonly serves to denote negative valence (“His scores were low”).

Yet, other conceptual frameworks for pitch mapping are conceivable. Mappings may originate in the experience of sound production – in qualities believed to be associated with objects producing “low” or “high” pitch. Primary among such relationships is the correlation of pitch and physical size, particularly body size. Cross-cultural studies of sound symbolism in language suggest an association of physical magnitude with pitch, such that larger size and mass, as well as secondary qualities associated with bodily magnitude, like slower pace and dominance, are related to lower-pitched tones in speech and in other vocal utterances (Berlin, 1994; Nichols, 1971; Nuckolls, 1999; Tsur, 1992). The association of pitch and body magnitude, and its extensions to secondary qualities, particularly emotional ones, may have a strong biological, evolutionary basis. Morton (1977, 1982) observed a widespread tendency among mammals and birds to use lower-pitched, rough voice when hostile and aggressive, and higher-pitched, tone-like voice when friendly or submissive. These acoustic features, pitch height in particular, are also correlated across species (though not necessarily within species) with body size, since body size usually correlates with the length of the vocal tracts, which largely determines pitch. In his “motivational–structural rules,” Morton suggests that since larger body size provides a clear advantage in a physical confrontation, animals in a confrontational mode would try to appear larger, not only visually (e.g., erecting their feathers, raising their tail) but by adopting the vocal cues for a large body size, such as low fundamental frequency (F0). Similarly, a submissive or friendly animal would “shrink” its apparent body size to appear less threatening, using both visual cues (e.g., flattening its tail or feathers) and auditory ones, particularly a high-pitched, purer voice (see also Chuenwattanapranithi, Xu, Thipakorn, & Maneewongvatana, 2008; Ohala, 1982, 1994; Scherer, 2003, 2004). Thus, the pitch-size correspondence may serve as a natural, biologically-based hub for a host of pitch metaphors, including those associating high and low pitch with a variety of emotions, as well as metaphors related to rank and social status. Importantly, pitch metaphors based upon the pitch-size correspondence may conflict with those based on the pitch-height analogy discussed earlier, since in the pitch-size association high pitch is “less” (smaller) while in the pitch-height association it is “more” (spatially higher).

Another source of mappings of pitch may stem from the notions of “intensity” or “tension.” More force and greater tension of a surface tend to cause higher sounds. Abstract mappings of pitch may in turn stem from a generalized notion of “intensity,” which would map higher pitch to higher intensity in other domains, such as brighter colors or faster pace (see Marks, 2004; Stevens, 1975 for overview of cross-modal intensity analogies; Eitan, 2007; Eitan & Gra-not, 2007, for surveys of musically-related studies of the topic). More generally, high and low pitch may be related to other polar dimensions via a generalized magnitude representation or shared mechanisms for processing magnitude relationships (Cohen Kadosh, Lammertyn, & Izard, 2008; Walsh, 2003), such that higher pitch is associated with higher magnitude poles in other domains. Note that (like mappings of pitch and spatial height, discussed above) such an account of pitch mappings seems to conflict with the relationship of high pitch and small size (i.e., lower magnitude) referred to above. We shall return to this dichotomy in the general discussion.

The conceptual metaphorical frameworks underlying pitch mapping may not always have a direct, explicit expression in the vocabulary. Their presence may be latent, and indirectly effect cognitive operations concerning auditory pitch even in the lack of explicit verbal expressions (Marks, 1996; Seitz, 2005). One empirically testable hypothesis stemming from a notion of basic latent mappings is that Western subjects would tend to agree on the “correct” application of non-Western pitch metaphors (such as those presented above), and that their mappings will be congruent with the original (i.e., non-Western or historical) application of those metaphors, which may be completely unfamiliar to them. This may be the case if, as suggested above, pitch metaphors, while culturally diverse, may be based upon basic underlying mappings, stemming from bodily-based inter-modal interactions with the physical environment. The experiments presented here aim at investigating such latent mappings and tracing their underlying dimensions.

1.1. General design and hypotheses

We report here four experiments, investigating the cognitive structure of pitch metaphors and the relevance of such metaphors to music listening. Experiment 1 is an exploratory conceptual investigation, in which Western participants matched 29 antonym metaphor pairs to two pitch poles (high–low). The antonyms were mostly collected from cross-cultural data and from results of previous perceptual and cognitive studies. Experiment 2 investigated how these antonyms are applied to an actual musical context – two segments contrasting in pitch register from Beethoven’s piano sonata, opus 111. In Experiment 3, participants applied the same antonyms to the poles of spatial verticality (high–low), rather than to auditory pitch. Finally, in Experiment 4, these antonyms are evaluated with respect to six basic dimensions, including size, mass, quantity, valence and intensity, besides pitch height and spatial height.

Several questions underlay these experiments. First, we examine whether pitch mappings not generally used in participants’ language and culture are nevertheless consistently applied. Specifically, we examine how Western subjects apply non-Western pitch mappings (Experiments 1 and 2). Second, we examine whether pitch mappings applied in an abstract conceptual task (Experiment 1) are similarly applied by listeners to actual music (Experiment 2). Third, we examine to what extent mappings of “high”
and “low” pitch correlate with evaluations of other dimensions, including spatial verticality, size and intensity (Experiments 3 and 4).

2. Experiment 1: metaphorical mappings of auditory pitch

2.1. Introduction

We begin exploring the cognitive mappings of auditory pitch using two simple conceptual tasks. Participants (native Hebrew speakers) were presented with 29 antonym pairs. The majority of these pairs were derived from pitch metaphors unfamiliar to the participants. They included metaphors derived from non-Western or historical usage, suggested by cross-cultural studies of sound symbolism in language, by cognitive and perceptual experiments, or by biological cross-species research, as surveyed above. Participants were asked to indicate what term in each antonym pair would be more appropriate as a metaphor for high pitch, and what term – for low pitch (task 1). In addition, they rated to what degree the antonym as a whole is appropriate as a description of (or a metaphor for) auditory pitch (task 2).

In this experiment, we ask Western participants to apply to the pitch domain terms which are not habitually used in their own language to describe that domain. We assume that decisive agreement on the application of such terms would provide support to the notion of latent pitch metaphors raised above. Such support will be particularly strong if a consensus on the application of an antonym to pitch (i.e., which term goes with high pitch, and which term goes with low) would be found even when the pair as a whole is judged a poor metaphor for pitch in task 2.

2.2. Method

2.2.1. Participants

Sixty-three Tel Aviv University students (27 males, 36 females; mean age = 26.3, SD = 8.13), native Hebrew speakers, 28 of them musically trained (>7 years of formal musical studies). Musically trained participants had on average more than 9 years of formal music training (mean = 9.18, SD = 3.7), while musically untrained participants had on average less than one year of musical lessons (mean = 0.84, SD = 1.79). Participants were paid approximately 8$ for a 45 min session.

2.2.2. Materials

In the pre-test (see “procedure” below), 26 pairs of pitches, varying in register, duration, timbre, loudness and articulation, 13 rising and 13 falling (intervals within pairs varied from one semitone to three octaves), were created using instrumental samples of the Sibelius 2 music software. Sounds were recorded on a CD and presented through loudspeakers.

In the main test, the only material used was a paper questionnaire that participants filled in manually. The questionnaire contained a list of antonym pairs in random order. No auditory stimuli were used.

2.2.3. Procedure

The session started with a pre-test, administered to distinguish participants with better and poorer aptitude for pitch direction discrimination so that results of the two groups in the main experiment could be compared. Participants heard the 26 pairs of pitches, presented in random order, with approximately 5 s between pairs, and were asked to indicate which pitch in each pair is higher (or lower, depending on the experiment group).

The main test took part after the pre-test, following a 5 min break. In part 1 of the main test, participants received a list of 29 pairs of antonyms, and were asked to note which term in each pair is a better description or metaphor for high pitch (task 1), and mark on a scale of 1–5 how well does each pair of antonyms as a whole serve as a metaphor for auditory pitch (task 2). Antonyms were presented in four different random orderings, while the order of antonyms within pairs was counterbalanced among participants. Participants were given 10–15 min to complete the tasks, and a 5-min break before proceeding to part 2.

In part 2, participants received the same list of antonyms (presented in different random orderings, but preserving ordering within pairs), and were asked to note which term in each antonym pair is a better description or metaphor for low pitch (task 1). Task 2 was identical to that in part 1. The order of parts 1 (high pitch) and 2 (low pitch) were counterbalanced among participants. The experiment was administered in groups of 10–15 people.

2.3. Results

2.3.1. Matching task

Table 1 presents an overview of the main results. The table indicates the proportion of participants that chose the listed pole of an antonym to represent high pitch (H, column 2) or low pitch (L, column 3), and indicates whether this proportion significantly differs from 50%. Chi² tests were done for each antonym and for the High and Low condition separately. An FDR correction (Benjamini & Hochberg, 1995) was made for multiple testing. Due to missing data points (answers left blank) N ranged between 56 and 63 (df = 1).

Generally, there was a substantial agreement among participants as to what pole of an antonym fits high or low pitch. The median proportion of participants that chose the listed poles to represent high pitch was .86, while the upper quartile was .95. In contrast, the median proportion of participants that chose these poles to represent low pitch was .18, and the lower quartile was only .07. Indeed (as can be seen in Table 1), for almost all antonyms, the agreement between participants was significantly above chance level, and for many it neared consensus, as agreement was >.90 for 13 of the metaphors for high pitch, and for 10 metaphors for low pitch. Note that these highly consensual terms included all metaphors derived from non-Western or historical usage, except strong–weak. Only two antonym pairs – far–near and hard–soft – were not significantly associated with pitch height. For all other antonyms, at least one pole (but generally both) was chosen with significant agreement. The strongest consensus was
but high pitch is not empty. These asymmetries form intu-
but low pitch is not soft and ugly, while low pitch is full,
for low pitch. In other words, high pitch is loud and pretty,
the proportion of participants who chose a
marked in
interaction between condition (high vs. low) and
was done for each antonym separately to examine the
An FDR correction was made for multiple
interaction between participants' pre-test scores and
The analysis compared the
groups whose pre-test scores were above
90% correct (N = 34) with those who scored 90% or lower
(N = 29). After FDR correction for multiple testing, no
significan
differences were found in the response distributions
of the two groups.

2.3.1.4. Appropriateness ratings. Importantly, in contrast to
the decisive agreement among participants concerning the
"proper" application of most antonyms to pitch (task 1),
the majority of the antonyms were rated as relatively
appropriate, with only one average rating (for thin–thick)
exceeding 3.5 (on a 1–5 scale). The median rating was 2.8
and 2.9 for the high and low pitch conditions, respectively.
The upper quartiles of the ratings were only 3.2 and 3.1
for the high and low pitch conditions, respectively, while the
lower quartile of the ratings was 2.6 for both conditions.

The association between average ratings for an anto-
nym pair (task 2) and the consensus between participants
regarding its application to pitch (task 1) was moderate:
the correlation between average rating and the proportion
of participants choosing the high pole is .58 for the high
pitch condition and .56 for the low pitch condition.
This moderate correlation is mostly due to the restricted varia-
tion in ratings of the antonyms; notably, even antonyms
that show high consensus (for example, the historical or
non-Western metaphors sharp–heavy and grandmother–
daughter) were often given a medium to low appropri-
ateness rating. Other antonyms were, however, moderately
rated, but showed no consensus. This was the case, for
example, for weak–strong and hard–soft.

As mentioned, participants rated the appropriateness of
each antonym twice. Though the two rating tasks were
identical ("rate how appropriate the antonym, pair as a
whole is as a description or metaphor for pitch"), one of
them followed task 1 for high pitch (which pair member
is a better description/metaphor for high pitch"), while
the other followed task 1 for low pitch. Generally, ratings
in these two contexts corresponded well (r = .85). Never-
theless, a few ratings showed a significant asymmetry, as
assessed by a paired samples T-test (as in Table 1): the ant-
onyms alert–sleepy and tense–relaxed were rated significa-
cantly higher in the high pitch context, while the
antonyms light–heavy and empty–full were rated higher
in the low pitch context.

### Table 1

Overview of Experiment 1 results, including proportion of participants choosing the high pole in the high (H) or low (L) condition, and appropriateness ratings for the high (H) and low (L) condition. Crosses indicate significant differences between the high and low condition (As.).

<table>
<thead>
<tr>
<th>Consensus</th>
<th>Appropriateness</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>H</td>
</tr>
<tr>
<td>Active–passive</td>
<td>.87d</td>
</tr>
<tr>
<td>Alert–sleepy</td>
<td>1.0a</td>
</tr>
<tr>
<td>Beginning–end</td>
<td>.66a</td>
</tr>
<tr>
<td>Cold–hot</td>
<td>.72d</td>
</tr>
<tr>
<td>Empty–full</td>
<td>.62</td>
</tr>
<tr>
<td>Far–near</td>
<td>.58</td>
</tr>
<tr>
<td>Fast–slow</td>
<td>.95d</td>
</tr>
<tr>
<td>Feminine–masculine</td>
<td>.90d</td>
</tr>
<tr>
<td>Fool–wise</td>
<td>.51</td>
</tr>
<tr>
<td>Grdauh–Grdma</td>
<td>.93</td>
</tr>
<tr>
<td>Happy–sad</td>
<td>.95d</td>
</tr>
<tr>
<td>Hard–soft</td>
<td>.57</td>
</tr>
<tr>
<td>Light–dark</td>
<td>.95d</td>
</tr>
<tr>
<td>Light–heavy</td>
<td>.95d</td>
</tr>
<tr>
<td>Little–much</td>
<td>.57</td>
</tr>
<tr>
<td>Loud–soft</td>
<td>.76d</td>
</tr>
<tr>
<td>Pleasant–unpleasant</td>
<td>.53</td>
</tr>
<tr>
<td>Pretty–ugly</td>
<td>.86</td>
</tr>
<tr>
<td>Right–left</td>
<td>.75</td>
</tr>
<tr>
<td>Rough–smooth</td>
<td>.85d</td>
</tr>
<tr>
<td>Sharp–blunt</td>
<td>.97d</td>
</tr>
<tr>
<td>Sharp–heavy</td>
<td>.98d</td>
</tr>
<tr>
<td>Small–large</td>
<td>.90d</td>
</tr>
<tr>
<td>Sparse–dense</td>
<td>.72</td>
</tr>
<tr>
<td>Strong–weak</td>
<td>.64d</td>
</tr>
<tr>
<td>Summer–winter</td>
<td>.80d</td>
</tr>
<tr>
<td>Tense–relaxed</td>
<td>.90d</td>
</tr>
<tr>
<td>Thin–thick</td>
<td>.98d</td>
</tr>
<tr>
<td>Young–old</td>
<td>.97d</td>
</tr>
</tbody>
</table>

Proportion of participants choosing high or low is significantly different from 50%.

| a | p < .03.  |
| b | p < .01.  |
| c | p < .001. |
| d | p < .0001.

reached for alert–sleepy, sharp–heavy, thin–thick, young–
old, sharp–blunt and fast–slow.

### 2.3.1. High–low asymmetries.

Chi² tests (df = 1) were done for each antonym separately to examine the interaction between condition (high vs. low) and distribution of responses, the null hypothesis being that participants would make opposite choices for high and low pitch. An FDR correction was made for multiple testing. Due to missing data points N ranged between 56 and 63 (df = 1). After FDR correction, none of the differences between the groups remained significant, suggesting that musical expertise does not play a major role in listeners’ ratings of pitch metaphors.

2.3.1.2. Musicians vs. non-musicians. A series of Chi² tests were done for each antonym separately to examine the interaction between musical training and the distribution of responses. An FDR correction was made for multiple testing. Due to missing data points N ranged between 56 and 63 (df = 1). After FDR correction, none of the differences between the groups remained significant, suggesting that musical expertise does not play a major role in listeners’ ratings of pitch metaphors.

2.3.1.3. Poor vs. good pre-test performers. Another series of
Chi² tests was done for each antonym to examine the interaction between participants’ pre-test scores and the distribution of their responses. The analysis compared the group of participants whose pre-test scores were above 90% correct (N = 34) with those who scored 90% or lower (N = 29). After FDR correction for multiple testing, no significant differences were found in the response distributions of the two groups.

2.3.1.4. Appropriateness ratings. Importantly, in contrast to
the decisive agreement among participants concerning the
"proper" application of most antonyms to pitch (task 1),
the majority of the antonyms were rated as relatively
appropriate, with only one average rating (for thin–thick)
exceeding 3.5 (on a 1–5 scale). The median rating was 2.8
and 2.9 for the high and low pitch conditions, respectively.
The upper quartiles of the ratings were only 3.2 and 3.1
for the high and low pitch conditions, respectively, while the
lower quartile of the ratings was 2.6 for both conditions.

The association between average ratings for an anto-
nym pair (task 2) and the consensus between participants
regarding its application to pitch (task 1) was moderate:
the correlation between average rating and the proportion
of participants choosing the high pole is .58 for the high
pitch condition and .56 for the low pitch condition.
This moderate correlation is mostly due to the restricted varia-
tion in ratings of the antonyms; notably, even antonyms
that show high consensus (for example, the historical or
non-Western metaphors sharp–heavy and grandmother–
daughter) were often given a medium to low appropri-
ateness rating. Other antonyms were, however, moderately
rated, but showed no consensus. This was the case, for
example, for weak–strong and hard–soft.

As mentioned, participants rated the appropriateness of
each antonym twice. Though the two rating tasks were
identical ("rate how appropriate the antonym, pair as a
whole is as a description or metaphor for pitch"), one of
them followed task 1 for high pitch (which pair member
is a better description/metaphor for high pitch"), while
the other followed task 1 for low pitch. Generally, ratings
in these two contexts corresponded well (r = .85). Never-
theless, a few ratings showed a significant asymmetry, as
assessed by a paired samples T-test (as in Table 1): the ant-
onyms alert–sleepy and tense–relaxed were rated significa-
cantly higher in the high pitch context, while the
antonyms light–heavy and empty–full were rated higher
in the low pitch context.
2.4. Discussion

Our first experiment complements previous studies (e.g., Marks, 1987; Walker & Smith, 1984) by demonstrating a strong consensus on the application of diverse metaphors, across sensory modes and conceptual realms, to the domain of auditory pitch. We have seen that for a vast majority of participants, regardless of musical training or pitch perception aptitude, high pitches are small, thin, sharp, smooth, bright, fast, alert, tense, young, and female (to name just the most consensual metaphors), while low pitches are large, thick, heavy, blunt, rough, slow, sleepy, old and male. Pitch, then, possesses a thick web of cross-domain relationships, intuitively agreed upon by most participants.

Moreover, consensus on the application of non-Western and historical metaphors was particularly high. Participants agreed on the “proper” application of these metaphors, and matched them with high and low pitch in accordance with their original applications, of which our Western participants had no previous knowledge. Furthermore, though participants rated the antonym pairs presented to them as only moderately (and sometimes poorly) appropriate as metaphors for pitch, they still agreed strongly on how these metaphors should be applied. Together, these two findings – the consensus on the application of non-Western metaphors and the strong agreement on the application of even poorly-rated antonyms – strengthen the notion of “latent” cross-domain mappings, widely understood and agreed upon despite their unfamiliarity.

Experiment 1 was purely conceptual, as participants matched verbal antonyms rather than actual auditory stimuli. In previous experiments concerning pitch metaphors (e.g., Smith & Walker, 1984, 1986) simple auditory stimuli, such as pure tones, were presented. The question addressed in Experiment 2 is whether metaphorical mappings of pitch observed in Experiment 1 and in previous research also apply to complex auditory stimuli. In particular, we ask whether these mappings apply to a musical context – an issue of particular interest, given the rich metaphor-laden discourse in Western music theory and criticism.

3. Experiment 2: metaphorical mappings of auditory pitch in a musical context

3.1. Introduction

In Experiment 2, participants listened to two musical segments which contrast in pitch range (high vs. low), but are similar in other respects. The stimuli were taken from the variations movement of Beethoven’s piano sonata, opus 111 (MM. 65–72, and 73–80) of Beethoven’s piano sonata, opus 111. For each segment, participants were presented with 35 antonym pairs, among them all 29 pairs used in Experiment 1. To enable more nuanced responses to the complex musical stimuli, the participants’ task differed from that in Experiment 1. Rather than selecting between two antonyms, they were presented with a bipolar scale for each antonym pair, and could thus indicate to what degree each term applies to each of the two musical segments.

Clearly, musical variables other than pitch range, such as the specific musical instrument used (piano), may affect participants’ ratings of different metaphors. Hence, we did not expect an exact correspondence between results of Experiments 1 and 2. Yet, compatible responses in the two experiments would indicate that the abstract mappings of Experiment 1 do have ecological validity, and are applied, on the whole, even to Beethoven’s late music – perhaps one of the most complex, richest auditory artefacts one may find.

3.2. Method

3.2.1. Participants

Sixty-three Tel Aviv University students (28 males, 35 females; mean age = 29.9, SD = 12.98), native Hebrew speakers, 36 of them musically trained (>7 years of formal musical studies) participated in the experiment. Musically trained participants had on average more than 9 years of formal music training (mean = 9.86, SD = 3.68), while musically untrained participants had on average 1 year of formal music training (mean = 1.00, SD = 1.33). Participants were paid approximately $6 for a 30 min session. None of the participants participated in any of the other experiments.

3.2.2. Materials

Two segments from the 2nd movement (Arietta) of Beethoven’s piano sonata, opus 111 (MM. 65–72, and 73–80) were used. These segments function analogously in a variation form. They share an underlying structure, and are similar in rhythm and texture, but contrast in pitch register (high vs. low). The mean pitch distance between the upper lines of the two segments is 2 octaves (24 semitones). The mean frequency in the lower segment was 228 Hz, ranging from C2 (65 Hz), its bass note, to G4 (392 Hz), the top note of its upper line. The mean frequency of the higher segment was 1088 Hz, ranging from G#4 (415 Hz) in its lower line to A6 (1760 Hz) in its upper line. The segments’ duration was 37 (high) and 40 (low) seconds. Both segments were taken from Daniel Barenboim’s recording of Beethoven’s piano sonatas (EMI Classical CZSS79122). For the sake of ecologic validity, no loudness normalization or other modification was made to the recordings.

3.2.3. Procedure

Participants were tested in groups of 10–15. They received a list of 35 antonyms (including all antonyms used in Experiment 1, and six additional ones; see the bottom rows of Table 2). Different random orderings of the 35 pairs were used, and the order of words within pairs was randomized and counterbalanced across participants. Participants listened once to each of the two musical segments (presented in counterbalanced order to different participant groups). Immediately following each listening, they marked on a 1–5 bipolar scale how appropriate were the terms as a metaphorical description of the music (1 – the term to the right is very appropriate; 5 – the term to the left is very appropriate). There was no time limit for responses.
### Table 2
Overview of Experiment 2 results. Listed are the ratings for the high-pitched (H) and low-pitched (L) music, the difference between these ratings (Δ), and its significance level. Antonyms showing a significant interaction between experiment (1 or 2) and response distribution are indicated in the rightmost column (Diff. Exp. 1).

<table>
<thead>
<tr>
<th>Antonym</th>
<th>H</th>
<th>L</th>
<th>Δ</th>
<th>p</th>
<th>Diff. Exp. 1</th>
</tr>
</thead>
<tbody>
<tr>
<td>Active–passive</td>
<td>3.4</td>
<td>3</td>
<td>0.4</td>
<td>×</td>
<td></td>
</tr>
<tr>
<td>Alert–sleepy</td>
<td>3.8</td>
<td>2.6</td>
<td>1.2</td>
<td>d</td>
<td></td>
</tr>
<tr>
<td>Beginning–end</td>
<td>2.8</td>
<td>1.6</td>
<td>1.2</td>
<td>d</td>
<td></td>
</tr>
<tr>
<td>Cold–hot</td>
<td>2.1</td>
<td>1.6</td>
<td>0.5</td>
<td>d</td>
<td></td>
</tr>
<tr>
<td>Empty–full</td>
<td>2.2</td>
<td>1.2</td>
<td>1</td>
<td>d</td>
<td></td>
</tr>
<tr>
<td>Far–near</td>
<td>3</td>
<td>3</td>
<td>0</td>
<td>d</td>
<td></td>
</tr>
<tr>
<td>Fast–slow</td>
<td>3.7</td>
<td>2.4</td>
<td>1.3</td>
<td>d</td>
<td></td>
</tr>
<tr>
<td>Feminine–masculine</td>
<td>3.1</td>
<td>1.1</td>
<td>2</td>
<td>d</td>
<td></td>
</tr>
<tr>
<td>Fool–wise</td>
<td>1.6</td>
<td>1.4</td>
<td>0.2</td>
<td>d</td>
<td></td>
</tr>
<tr>
<td>Grduaugh–Grdma</td>
<td>3.2</td>
<td>1.1</td>
<td>2.1</td>
<td>d</td>
<td></td>
</tr>
<tr>
<td>Happy–sad</td>
<td>3.7</td>
<td>2.4</td>
<td>1.3</td>
<td>d</td>
<td></td>
</tr>
<tr>
<td>Hard–soft</td>
<td>2.1</td>
<td>3.1</td>
<td>1</td>
<td>d</td>
<td></td>
</tr>
<tr>
<td>Light–dark</td>
<td>3.4</td>
<td>0.8</td>
<td>2.6</td>
<td>d</td>
<td>×</td>
</tr>
<tr>
<td>Light–heavy</td>
<td>3.5</td>
<td>1</td>
<td>2.5</td>
<td>d</td>
<td></td>
</tr>
<tr>
<td>Little–much</td>
<td>2.4</td>
<td>1.7</td>
<td>0.7</td>
<td>a</td>
<td></td>
</tr>
<tr>
<td>Loud–soft</td>
<td>2.3</td>
<td>3.1</td>
<td>−0.8</td>
<td>d</td>
<td>×</td>
</tr>
<tr>
<td>Pleasant–unpleasant</td>
<td>4</td>
<td>3.3</td>
<td>0.7</td>
<td>b</td>
<td></td>
</tr>
<tr>
<td>Pretty–ugly</td>
<td>4.2</td>
<td>3.3</td>
<td>0.9</td>
<td>d</td>
<td></td>
</tr>
<tr>
<td>Right–left</td>
<td>3.4</td>
<td>2.9</td>
<td>0.5</td>
<td>d</td>
<td></td>
</tr>
<tr>
<td>Rough–smooth</td>
<td>1.9</td>
<td>3.5</td>
<td>−1.6</td>
<td>d</td>
<td></td>
</tr>
<tr>
<td>Sharp–blunt</td>
<td>3.8</td>
<td>2</td>
<td>1.8</td>
<td>d</td>
<td></td>
</tr>
<tr>
<td>Sharp–heavy</td>
<td>3.7</td>
<td>1.9</td>
<td>1.8</td>
<td>d</td>
<td></td>
</tr>
<tr>
<td>Small–large</td>
<td>3.1</td>
<td>1.3</td>
<td>1.8</td>
<td>d</td>
<td></td>
</tr>
<tr>
<td>Sparse–dense</td>
<td>2.6</td>
<td>3.1</td>
<td>0.5</td>
<td>d</td>
<td></td>
</tr>
<tr>
<td>Strong–weak</td>
<td>2.5</td>
<td>3.4</td>
<td>−0.9</td>
<td>b</td>
<td>×</td>
</tr>
<tr>
<td>Summer–winter</td>
<td>3.9</td>
<td>2</td>
<td>1</td>
<td>d</td>
<td></td>
</tr>
<tr>
<td>Tense–relaxed</td>
<td>2.8</td>
<td>3.4</td>
<td>−0.6</td>
<td>d</td>
<td>×</td>
</tr>
<tr>
<td>Thin–thick</td>
<td>3.2</td>
<td>0.8</td>
<td>2.4</td>
<td>d</td>
<td></td>
</tr>
<tr>
<td>Young–old</td>
<td>3.2</td>
<td>1.1</td>
<td>2.1</td>
<td>d</td>
<td></td>
</tr>
<tr>
<td>Cause–result</td>
<td>2</td>
<td>1.9</td>
<td>0.1</td>
<td>d</td>
<td></td>
</tr>
<tr>
<td>Clean–dirty</td>
<td>4.4</td>
<td>2.9</td>
<td>1.5</td>
<td>d</td>
<td></td>
</tr>
<tr>
<td>Follow–crocodile</td>
<td>2.3</td>
<td>1.4</td>
<td>0.9</td>
<td>a</td>
<td></td>
</tr>
<tr>
<td>Good–bad</td>
<td>4</td>
<td>3.1</td>
<td>0.9</td>
<td>d</td>
<td></td>
</tr>
<tr>
<td>Sane–crazy</td>
<td>2.6</td>
<td>2.4</td>
<td>0.2</td>
<td>d</td>
<td></td>
</tr>
<tr>
<td>Sweet–sour</td>
<td>4.1</td>
<td>2.6</td>
<td>1.5</td>
<td>d</td>
<td></td>
</tr>
</tbody>
</table>

a p < .05.
b p < .01.
c p < .001.
d p < .001.

### 3.3. Results

The average ratings of the antonyms for the high- and low-pitched musical fragments are listed in Table 2. Higher ratings (on a 1–5 scale) indicate a stronger tendency to prefer the term listed first in column 1, while lower ratings indicate a tendency to prefer its antonym. Thus, for instance, “thin” (row 3) was rated as relatively appropriate for the high-pitched segment (3.24 average rating), while “thick” was rated as appropriate for the low-pitched segment (0.75 average rating).

A paired samples T-test indicates whether the differences between the ratings of antonyms for high and low musical fragments significantly deviate from 0 (N varied between 58 and 63, due to some missing data points). Since the main distinction between the two segments was their pitch range, we assume that these rating differences are an indication for an effect of pitch range on the application of these metaphors to music. For most antonyms (27 out of 35), the difference was significant after FDR correction (p < .03) and for many it was highly significant (p < .0001). As in Experiment 1, terms significantly associated with pitch include all but one (crazy–sane) of the metaphor pairs derived from ethnographical or historical sources, including “exotic” metaphors such as “crocodile” or “grandmother.”

The 10 antonyms that showed the most significant differences between the high and low conditions were thin–thick, light–dark, light–heavy, sharp–heavy, young–old, granddaughter–grandmother, feminine–masculine, rough–smooth, sharp–blunt and large–small (in descending order of difference). The antonyms that showed no significant difference were cause–result, far–near, crazy–sane, fool–wise, right–left, active–passive, cold–hot, tense–relaxed (in increasing order of difference).

#### 3.3.1. Effects of musical training

Using independent T-tests with FDR correction for multiple testing, we compared the mean ratings by musicians and non-musicians for all metaphors applied to the high and low music samples. Only two significant differences were found: non-musicians’ ratings of the metaphor “loud” in the low segment were significantly higher than musicians’ ratings (means = 3.52, SD = 0.98 for non-musicians vs. mean = 2.84, SD = 0.86 for musicians; p < .05), and the musicians’ ratings of “right” (direction) for the high pitch segment were significantly higher than non-musicians’ ratings (means = 2.85, SD = 1.13 for non-musicians vs. mean = 3.76, SD = 1.10 for musicians; p < .05).

#### 3.3.2. Comparing Experiments 1 and 2

To compare the results of Experiment 2 with Experiment 1, the data of Experiment 2 was converted, for each participant, into a binary preference coding, by subtracting the rating of an antonym in the high-pitched music from its rating in the low-pitched music, and coding the result as 1 (for a positive result) or 0 (for a negative result). For instance, given a scale of 5 (“light” is highly appropriate as a metaphor for the relevant music) to 1 (“dark” is highly appropriate for the given music), 1 is used for a participant who for the light–dark antonym marked 5 in the high-pitched music and 3 in the low-pitched music (suggesting that he associated “light” with higher pitch), while datum for a participant that, for that antonym, marked 1 in the high-pitched music and 3 in the low-pitched music is coded 0 (associating “dark” with higher pitch). Equal ratings for high and low music were excluded from the data for this test. In Experiment 1, we considered task 1 data in both the low pitch and the high pitch conditions, and excluded results in which data (for a specific participant) for high and low was inconsistent, e.g., as when a participant indicated that “dark” is appropriate as a metaphor for both high and low pitch.

Chi² tests were conducted for each antonym separately to examine the interaction between experiment and the distribution of responses. An FDR correction was

---

1 The difference between musically trained and untrained participants with regard to the association of pitch and laterality is in accord with previous research (e.g., Eitan & Granot, 2006; Lidji et al., 2007; Stewart, Walsh, & Frith, 2004).
conducted for multiple testing. Due to some missing data, \( N \) varied between 76 and 116 (df = 1). The last column of Table 2 indicates the antonyms for which a significant difference was found between Experiments 1 and 2 (\( p < .03 \)). The strongest interactions concerned the antonyms tense–relaxed and soft–loud (\( \chi^2 = 40, p < .0001; \chi^2 = 30, p < .0001 \)). These were followed by the antonyms hard–soft, strong–weak, alert–sleepy and pleasant–unpleasant (\( \chi^2 = 15, p < .0001; \chi^2 = 14, p < .0001; \chi^2 = 11, p < .01; \chi^2 = 11, p < .01 \)). For hard–soft, strong–weak, and pleasant–unpleasant the association between pitch and the poles of the antonyms was more decisive in Experiment 1 than in Experiment 2. The only opposite association concerned the antonym loud–soft. In Experiment 2, low was associated with loud, while in Experiment 1, high was associated with loud.

3.3.3. Factor analysis

To highlight groups of antonyms that reveal correlated responses, the differences between ratings of the antonyms for the high and low music in Experiment 2 were subjected to a principle component analysis. Only the 27 antonyms that revealed significant differences between ratings for the high and low conditions were included in the analysis. The analysis reduced the 27 dimensions of antonyms to its main components, grouping correlated antonyms (\( N \) per antonym ranged between 56 and 62, due to some missing data points).

The principle component analysis resulted in eight components with an eigenvalue greater than 1. The four main components were selected for inclusion in a varimax rotation that aligns the components with variable clusters, thus increasing the interpretability of the analysis. These factors accounted for 17%, 11%, 17%, and 12% of the variance, respectively, which adds up to a total 58% of the variance. Table 3 shows the variables that correlated most strongly with each factor (\( r > .50, p < .0001 \)).

This list suggests a number of separate dimensions shaping pitch metaphors in musical contexts. Three of the four dimensions (F2, F3, and F4) correspond respectively with the dimensions of potency, valence and activity, well known from Osgood’s semantic differential analyses (Osgood, Suci, & Tannenbaum, 1957). The first dimension is a combination of brightness, sharpness, physical size, age and gender. The possible significance of this factor will be discussed below.

3.4. Discussion

Results of Experiment 2 suggest three main conclusions. First, listeners consistently apply varied cross-domain mappings to pitch not only in abstract or simple contexts, but within a complex musical context as well. Furthermore, these mappings largely correspond to those done in an abstract or rarified setting. Lastly, mappings group (based on the factor analysis) into a number of more basic dimensions, including the dimensions of activity, valence and potency, well known from research in diverse domains.

The considerable match between conceptual and music-related results is striking, given the fact that, in the musical context, many parameters, such as event density, texture, and instrumental timbre may interact with pitch register. For instance, the dense event rate in both high and low segments might have diminished the role of pitch register in rating “activity” and related metaphors, while the piano timbre might have emphasized pitch contrasts with regard to other antonyms, such as light–dark. Furthermore, the two musical segments, though they retained the same underlying structure, were not exact octave transfer of each other. Thus, the strong match between conceptual (Experiment 1) and musical (Experiment 2) metaphorical mappings highlights the seminal role of pitch register in shaping such mappings in a musical context.

Nevertheless, differences between results in the two experiments are noteworthy as well. As suggested above, some of these differences may be due to features of the music other than pitch register that influence the ratings of several antonyms. For example, event density and dynamics may have influenced ratings for active–passive and tense–relaxed more strongly than pitch, leading to small differences between the ratings in the low and high pitch conditions of Experiment 2. However, differences between the experiments may also be due to more general differences in connotations of musical pitch in actual musical circumstances, as distinguished from abstract conceptual judgments. For example, the stronger associations...
with pitch in the musical context for pleasant–unpleasant and strong–weak suggest that these particular mappings may be stronger or less ambiguous when an actual musical context is presented. Needless to say, further research, using different musical stimuli, is needed to support these conjectures.  

The factor analysis applied to results of Experiment 2 suggests that pitch metaphors tend to be correlated, such that the three dimensions often revealed in semantic differential analyses (e.g., Osgood et al., 1957) – evaluation, potency, and activity (EPA) – also underlie much of the co-variation concerning pitch metaphors.  

More intriguing is the remaining factor (F1), one of two factors accounting for the largest percentage of the variance, and containing the metaphors most consensually used in characterizing pitch. Indeed, all but one of the strongest variables in this factor can be related to mass or size, either directly (thin–thick, heavy in “sharp–heavy,” sharp–blunt) or indirectly (female–male, young–old), representing an “up is less” correlation of pitch and physical size or mass. Importantly, the above variables are experimentally related to pitch, rather than just products of cultural convention. By and large, objects producing lower pitch indeed tend to be thicker (and thus not as sharp), heavier, older (adults as compared to children), and male. Furthermore, the correlation between body size and mass and pitch register seem to affect animal behavior across species (Morton, 1977, 1982; see introduction), and its perception may thus be grounded in biological, innate predispositions.  

Yet, what makes this factor (F1) harder to interpret as a “mass” (or size/mass) dimension, is the visual lightness variable correlated with it. Cross-domain mapping of pitch and lightness has been observed in a number of studies in adults (Hubbard, 1996; Marks, 1987; Walker & Smith, 1984) and children as young as 2.5 years (Marks, Hammel, & Bornstein, 1987; Mondloch & Maurer, 2004). Unlike the mapping of pitch and physical size, mapping lightness and pitch is not directly related to experience: painting a surface white, rather than black, or exposing it to sunlight, would not change its pitch. Its early development suggests that it is neither linguistic in origin.  

Binding together visual lightness with concepts related to physical size and mass may suggest a “latent” cognitive dimension, not subsumed under any specific verbal concept, associated with auditory pitch. Results of Experiment 2 suggest that the possibility of such an underlying cross-modal, non-verbal representation for pitch should be considered.  

Experiments 3 and 4 continue the inquiry into basic dimensions underlying pitch metaphors in different ways. Experiment 3 focuses on a single related issue: do metaphorical mappings for pitch derive from those of spatial height? The verticality metaphor for pitch is central in Western languages and music notation. Hence, one may hypothesize that other metaphorical mappings for pitch are merely “second-order” mappings, which seem appropriate to pitch since they also apply to spatial verticality, habitually related to pitch. Our next experiment will examine this hypothesis.  

4. Experiment 3: metaphorical mappings of spatial elevation  

4.1. Introduction  

Verticity, the principal source domain for auditory pitch metaphors in many languages (including English, German and Japanese, as well as Hebrew, the language used by participants in the present experiment) is also applied to many other target domains. Lakoff and Johnson (1980a, 1980b) present several orientational metaphors (metaphors which structure concepts linearly, in analogy to spatial relationships) based on up–down relationships. Primary among these is “up is more,” mapping amount or quantity onto height (e.g., my income rose; the number of participants is very low). Perhaps derived from this basic quantitative mapping, different value-laden conceptual metaphors associate the preferred pole with higher elevation: “good is up” (e.g., “our quality of life is high”), “happy is up” (“I am down,” “cheer up”), and indirectly “control is up” (e.g., “he is under my control”) and “rational is up” (e.g., “the argument fell to the emotional level”).  

Importantly, several of these mappings have been shown to affect, and to be affected by, non-linguistic behavior, supporting the notion of its embodied nature. Thus, as Casasanto (2008) revealed, the vertical orientation of speakers’ gestures is consistent with that implied by the up–down schema, even when they do not use spatial vocabulary (e.g., speakers tended to move their hands up when uttering “my grades got better”). Correspondingly, when participants in Casasanto and Dijkstra (submitted for publication) were asked to recall positive or negative memories while moving marbles up or down, they retrieved more positive memories when moving marbles up, and more negative memories when moving them down. Moreover, both the speed of memory retrieval and that of bodily motion increased when movement direction was congruent with memory valence (up/positive, down/negative).  

Does the pitch-verticality mapping, then, represent another non-linguistic, embodied instantiation of the up/down metaphor, stemming from the same conceptual sources as other metaphorical mappings of the up–down polarity? Indeed, Cox (1999) suggested that the “up is more” metaphor is a principal source for the pitch-verticality mapping, among other things since rise in pitch corresponds with increased vocal effort and tension. The predominance of verticality mappings for musical pitch – both verbal and visual (as in music notation) – may suggest that other connotations of pitch, such as those exhibited in Experiments 1 and 2, are indirectly derived from metaphors for spatial verticality.  

In Experiment 3 we examine the hypothesis that the cross-domain mappings of high and low auditory pitch stem from those of spatial high and low in a simple way – by
asking participants to apply the antonyms presented in Experiment 1 not to auditory pitch, but to spatial elevation.

4.2. Method

4.2.1. Participants
Fifty-eight Tel Aviv University students (19 males, 39 females; mean age = 26.2, SD = 6.01), native Hebrew speakers. Ten of the participants were musically trained (>7 years of formal musical studies). Musically trained participants had on average 12 years of formal music training (mean = 12.0, SD = 3.83), while musically untrained participants had a little more than 1 year of musical training on average (mean = 1.35, SD = 2.17). Participants were paid approximately $8 for a 45 min session. None of the participants in Experiment 3 participated in any other experiment within this study.

4.2.2. Materials
Participants used paper forms identical to those used in Experiment 1, to which four antonyms related to sounds were added. In addition, participants heard two pairs of sampled sounds, differing in pitch, produced in pitch, produced in the Sibelius 2 music software. Pair 1 consisted of rectangle wave sounds, 100 and 500 Hz in frequency; pair 2 consisted of sinusoids, 200 and 600 Hz in frequency. All sounds were 500 ms in duration, and were presented to participants in a loudness level of approximately 60 dB-a. Sounds were recorded on a CD and presented through loudspeakers.

4.2.3. Procedure
The procedure of Experiment 1 (excluding the pre-test) was repeated, but participants were asked to apply metaphors to spatial height. Thus, in part 1 participants were asked which term in each pair of antonyms is a better description or metaphor for a high elevation (task 1), as well as rate on a numerical scale (1–5) how well does each pair of antonyms serve as a metaphor for spatial height (task 2). In part 2, participants were asked which term in each antonym pair is a better description or metaphor for low elevation (task 1); task 2 was identical to that in part 1. All antonyms featured in Experiment 1 were used. In addition, four pairs of verbal antonyms referring to sound, and two pairs of actual sounds differing in pitch (featured at the end of each part), were included in the questionnaire.

4.3. Results

4.3.1. Matching task
A series of Chi² tests were done for each antonym separately, to examine whether the distribution of responses differs between experiments (1 vs. 3). An FDR correction was conducted for multiple testing. Due to some missing data points N ranged between 111 and 121 (df = 1). Table 4 presents a comparison of results in Experiments 1 and 3, and indicates the antonyms for which a significant difference was found between the two experiments (p < .03).

As in Experiment 1, the consensus in choosing a pole of the antonyms to represent spatial high or low was generally high, with the majority of the proportions considerably differing from 50%.

Importantly, however, the proportions for spatial height deviate significantly from the proportions for pitch height for quite a few antonyms (16 out of 29). Several antonyms even showed opposite tendencies: small–large, female–male, empty–full, little–much, beginning–end, granddaughter–grandmother, and fool–wise were all matched in opposite ways for pitch height and spatial height. Thus, high pitch is small, empty, feminine, little, and beginning, while spatially high objects are large, full, masculine, wise, numerous in quantity and at the end. In other instances, the proportions significantly differ in magnitude. Consensus was significantly higher for pitch height with regard to thin–thick, young–old, light–dark, sharp–heavy, and granddaughter–grandma, suggesting that these antonyms fit pitch height more decidedly than spatial height. For three antonyms, strong–weak, pleasant–unpleasant and pretty–ugly (related to potency and valence), the consensus was in the same direction, but significantly higher for spatial height.

4.3.2. Appropriateness ratings
The appropriateness ratings of the antonyms for spatial height were generally moderate to low. The median rating was 2.8 for high and low spatial height, the upper quartile was 3.1 and 3.0 for high and low, respectively, and the lower quartile was 2.5 and 2.3 for high and low, respectively. In other words, the participants found the antonyms only moderately appropriate, despite the general consensus in choosing poles that correspond with spatially high or low positions.

Table 4 indicates the antonyms that were favored as significantly different in Experiment 1 and Experiment 3. In all these instances, the antonyms were rated significantly lower in Experiment 3 than in Experiment 1, except the antonyms soft–loud and small–large, which were rated higher in Experiment 3.

In short, not surprisingly, several of the antonyms were regarded as less appropriate as a metaphor for spatial height than for pitch height, despite the shared trend of generally moderate appropriateness ratings.

4.3.3. Auditory antonyms
Four verbal auditory antonyms and two actual pitches (high and low) were added in Experiment 3 to examine participants’ judgments of the mapping between these antonyms and spatial high and low. Table 4 (bottom lines) shows the consensus results and appropriateness ratings of these antonyms. The auditory antonyms denoting unambiguous pitch relations, soprano–bass and scream–sigh, showed the highest consensus and ratings. The other two antonyms, denoting high or low pitch less clearly, showed smaller consensus and lower ratings. The two high and low–pitched sounds were rated as good mappings of spatial height – indeed, consensus was between .92 and .96 and the ratings were between 3.85 and 3.97, higher than for most verbal antonyms.
Table 4
Results of Experiment 3 (spatial height) compared with those of Experiment 1 (pitch). In the “consensus” columns, proportions of Exp. 3 participants choosing the first antonym in each pair as a metaphor for high (H) and low (L) elevation are compared with proportions of Exp. 1 participants who chose that metaphor as a metaphor for high and low spatial elevation. In the “appropriateness” columns, average appropriateness ratings for each antonym pair in Exp. 1 (pitch) and 3 (spatial elevation) are compared.

<table>
<thead>
<tr>
<th>Consensus</th>
<th>Appropriateness</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Exp. 3 H</td>
</tr>
<tr>
<td>Active–passive</td>
<td>.86</td>
</tr>
<tr>
<td>Alert–sleepy</td>
<td>.91</td>
</tr>
<tr>
<td>Beginning–end</td>
<td>.24</td>
</tr>
<tr>
<td>Cold–hot</td>
<td>.51</td>
</tr>
<tr>
<td>Empty–full</td>
<td>.37</td>
</tr>
<tr>
<td>Far–near</td>
<td>.88</td>
</tr>
<tr>
<td>Fast–slow</td>
<td>.91</td>
</tr>
<tr>
<td>Feminine–masculine</td>
<td>.25</td>
</tr>
<tr>
<td>Fool–wise</td>
<td>.1</td>
</tr>
<tr>
<td>Grdau–Grdma</td>
<td>.39</td>
</tr>
<tr>
<td>Happy–sad</td>
<td>.84</td>
</tr>
<tr>
<td>Hard–soft</td>
<td>.6</td>
</tr>
<tr>
<td>Light–dark</td>
<td>.78</td>
</tr>
<tr>
<td>Light–heavy</td>
<td>.85</td>
</tr>
<tr>
<td>Little–much</td>
<td>.25</td>
</tr>
<tr>
<td>Loud–soft</td>
<td>.84</td>
</tr>
<tr>
<td>Pleasant–unpleasant</td>
<td>.88</td>
</tr>
<tr>
<td>Pretty–ugly</td>
<td>1</td>
</tr>
<tr>
<td>Right–left</td>
<td>.7</td>
</tr>
<tr>
<td>Rough–smooth</td>
<td>.83</td>
</tr>
<tr>
<td>Sharp–blunt</td>
<td>.89</td>
</tr>
<tr>
<td>Sharp–heavy</td>
<td>.86</td>
</tr>
<tr>
<td>Small–large</td>
<td>.18</td>
</tr>
<tr>
<td>Sparse–dense</td>
<td>.71</td>
</tr>
<tr>
<td>Strong–weak</td>
<td>.84</td>
</tr>
<tr>
<td>Summer–winter</td>
<td>.65</td>
</tr>
<tr>
<td>Tense–relaxed</td>
<td>.86</td>
</tr>
<tr>
<td>Thin–thick</td>
<td>.72</td>
</tr>
<tr>
<td>Young–old</td>
<td>.65</td>
</tr>
<tr>
<td>Scream–sigh</td>
<td>.98</td>
</tr>
<tr>
<td>Shout–whisper</td>
<td>.88</td>
</tr>
<tr>
<td>Sneeze–cough</td>
<td>.74</td>
</tr>
<tr>
<td>Soprano–bass</td>
<td>.98</td>
</tr>
<tr>
<td>Motif1</td>
<td>.95</td>
</tr>
<tr>
<td>Motif2</td>
<td>.96</td>
</tr>
</tbody>
</table>

a Antonyms showing a significant interaction (p < .03) between experiment (Exp. 1 vs. Exp. 3) and response distribution.

b Antonyms showing a significant difference (p < .03) between the appropriateness ratings in Exp. 1 and Exp. 3.

4.4. Discussion

The results of Experiment 3 and their relationship to those of the previous experiments present a complex and, in important ways, unique picture of the verticality metaphor for pitch and its relationship to other orientational metaphors, as described by Lakoff and Johnson (1980a, 1980b). On the one hand, participants’ matching of auditory antonyms and actual sounds with spatial high and low, and the high ratings given to these antonyms, clearly suggest a strong association of pitch and spatial verticality (thus corroborating results of previous experiments). However, results also show that participants associate pitch and spatial verticality with other dimensions, particularly size, gender, quantity, and evaluation, in different, often contrasting ways. Importantly, the conceptual mappings that serve as the very foundation of other verticality metaphors do not apply, or apply weakly, to auditory pitch. In particular, the “high is more” mapping, which (according to Lakoff and Johnson) directly or indirectly serves as the most important foundation of verticality metaphors (and, as mentioned, has also been proposed as the most important basis of the pitch–verticality mapping) does not apply to pitch. Rather, for pitch, “high is less” in several important ways: while spatially high objects are large and higher in quantity, high pitch is small, empty and little in quantity (in accord with Morton’s “motivational–structural rules,” 1977, 1982; see introduction above). Similarly, there is a very high consensus that higher pitches are thinner and sharper (i.e., of a smaller mass). In contrast with what might be expected, spatial elevation was also associated with thinner and sharper objects, although less strongly than pitch height.

Other significant differences between the connotations of pitch and spatial elevation apply to evaluative dimensions, which were suggested as foundational to other verticality metaphors, and affect both verbal behavior and bodily gesture (Casasanto, 2008; Casasanto & Dijkstra, submitted for publication; Meier & Robinson, 2005). Thus, spatially higher objects are more pleasant, while higher pitches – not necessarily so; spatially low objects are ugly, while low pitch is not; spatially low objects are weaker, while low pitch is not; and higher objects but lower pitches are wiser. Hence, it seems that “good is up,” “control is up,” and “rational is up” – the other basic conceptual mappings...
of verticality metaphors suggested by Lakoff and Johnson, also do not apply very well to pitch.

These results, as well as results of the factor analysis in Experiment 2, suggest that metaphors and imagery concerning pitch may stem from the interaction and conflict of several dimensions. In particular, higher pitch may often be associated not only (or even mainly) with superior spatial elevation and its connotations, but (as Morton, Ohala, and others have suggested) with lesser bodily size and its connotations. In Experiment 4, we further explore how these dimensions are associated with notions of auditory pitch and its production.

5. Experiment 4: multidimensional correlations

5.1. Introduction

Experiment 4 aims to further explore two issues suggested by results of the previous three experiments. First, it directly examines, using a within-subject design, the correlation of pitch metaphors with several other dimensions, including intensity, valence, size, quantity and spatial height. In the experiment, participants performed the two tasks (choosing between antonyms, and rating the appropriateness of antonym pairs) which had been used in Experiments 1 and 3, but applied them, in addition to pitch, to these other five dimensions. This investigation may help clarifying how the different (and sometimes conflicting) dimensions that may serve as hubs for diverse pitch metaphors – pitch and spatial height, pitch and size, pitch and intensity, or the more general pitch and magnitude mapping – act and interact.

Another issue examined in Experiment 4 is how participants construe the relationships of pitch metaphors with sound production – with perceived qualities of objects which produce high and low pitch. Several of the metaphors consensually applied to pitch – small/large, female/male, young/old – seem to derive from the qualities of sound-producing objects: females, young animals and humans, and smaller objects, indeed tend to produce higher pitches than males, adults, and larger objects. However, other consensual pitch metaphors, such as light/dark or summer/winter, do not clearly derive from such experienced correlations. In this experiment, rather than asking participants (as in Experiment 1) which term in each antonym pair would suit better as a metaphor for high/low pitch, we asked “Which object would tend to produce a higher-pitched/lower-pitched sound – one having quality (a) or one having quality b?” (questions referring to other dimensions were correspondingly phrased; see methods section below). Comparing responses to this question with those in Experiment 1 may thus shed some light on the relationship of pitch metaphors and auditory experience.

5.2. Method

5.2.1. Participants

Sixty Tel Aviv University students (27 males, 33 females; mean age = 24.9, SD = 3.94), native Hebrew speakers, 30 of them musically trained (>7 years of formal musical studies) participated in the experiment, and were paid approximately 16$ for two half an hour sessions. Musically trained participants had on average 13.4 years of musical training (SD = 5.13), while musically untrained participants had 1.26 years of musical training on average (SD = 1.89). None of the participants of Experiments 1, 2, and 3 took part in this experiment.

5.2.2. Materials

Participants filled out paper forms; no other equipment was used.

5.2.3. Procedure

The procedure was similar to those of Experiments 1 and 3, but applied to six different target domains: intensity, mass/size, valence, spatial height, pitch height (production), and quantity.

The experiment consisted of two sessions, each approximately 30 minutes long, a week apart. In each session, participants received three forms – one form for each question. Each form included 35 antonym pairs (identical to those in Experiment 2, and containing all 29 antonyms of Experiment 1). Two tasks were given for each form, analogous to those in Experiment 1. In task 1, participants were asked which term in each pair of antonyms is a better description of an object possessing a specified feature (see below). In task 2, they rated on a numerical scale (1–5) how well each pair of antonyms serves as a metaphor for the specified feature. As a control, opposing versions of each question in task 1 were counterbalanced among participants. For instance, half of the participants were asked (Q.2) “Which object would tend to produce a higher-pitched sound – one having quality (a) or one having quality (b)?” while the other half was asked which object would tend to produce a lower-pitched sound.

The six pairs of questions presented to the participants were:

Q.1 – Intensity:
I. Which object would tend to be more intense – one having quality (a) or one having quality (b)?
II. To what degree (1–5) is each pair associated with intensity?

Q.2 – Mass or size:
I. Which object would tend to be larger or heavier – one having quality (a) or one having quality (b)?
II. To what degree is each pair associated with mass or physical size?

Q.3 – Valence:
I. Which of each two qualities, (a) or (b), expresses a more positive evaluation (in any sense of the term)?
II. To what degree is each pair associated with evaluation?

Q.4 – Spatial height
I. Which object would tend to be located higher (in physical space) – one having quality (a) or one having quality (b)?
II. To what degree is each pair associated with spatial height?

Q.5 – Pitch height
I. Which object would tend to produce a higher-pitched sound – one having quality (a) or one having quality (b)?
II. To what degree is each pair associated with auditory pitch?

Q.6 – Quantity
I. What object would be associated with larger quantity or dimensions (in any sense, not necessarily physical) – one having quality (a) or one having quality (b)?
II. To what degree is each pair associated with quantity?

5.3. Results

5.3.1. Matching task
As noted, in this experiment participants matched the poles of each antonym with six different dimensions (task 1). To examine the associations between these dimensions, we computed pairwise Pearson correlation coefficients for task 1 (matching) responses for each pair of dimensions. This analysis presents the correlations between the choices of a particular pole of an antonym as “high” (or “positive”) or as “low” (or “negative”) in the different dimensions (intensity, pitch, spatial height, mass/size and valence). The consensus ranged from close to 1.0 to close to 0, depending on the antonym and dimension. For example, the consensus in choosing “thin” to represent “low” is high for mass/size (.93), but the consensus to chose thin to represent “low” in pitch is low (zero). All antonyms were included twice: once for the “high” condition of a dimension and once for the “low” condition (total N = 70).

Table 5 shows the results of this pairwise comparison for musicians and non-musicians separately. Specifically relevant are the correlations between pitch height and the other domains (column 1). As can be seen, the results for the two groups of participants are very similar. In particular, consensus results for pitch height correlate negatively with consensus results for size and quantity, indicating that low pitch is considered larger and more. The results correlate positively with consensus results for spatial height and valence, indicating that high pitch is considered spatially higher and more positive. Additionally, this table shows that antonym-choice was correlated for size and quantity and for valence and spatial height. Notably, choice of antonym for spatial height and size are negatively correlated, indicating that spatially high is in this experiment associated with lighter and smaller objects rather than with heavier and larger objects. Up is also not always more within the verticality domain.

To correct for inter-correlations between dimensions, Table 6 shows partial correlations between the variations in consensus across antonyms, as obtained for the different dimensions. Results are shown for musicians and non-musicians separately. Partial correlations give the unique correlation between two variables by adjusting for the correlation with all other variables. High partial correlations with pitch height were obtained only for intensity (positive correlation) and size (negative correlation). Notably, while the pairwise correlation of pitch with spatial height is strong, in partial correlations the association between these dimensions is considerably attenuated, especially for the musically trained participants, indicating that the correlation with spatial height might have been due in part to the association with a third variable. Similarly, the significant correlation of pitch and valence, present in the pairwise correlation, disappears in the partial correlation. In contrast, correlations of pitch and intensity, weak in the pairwise test, emerge as relatively strong in the partial correlations. The correlations with intensity seem to be relatively unique associations that do not also associate with the other dimensions.

<table>
<thead>
<tr>
<th>Musicians</th>
<th>Pitch</th>
<th>Intensity</th>
<th>Size</th>
<th>Valence</th>
<th>Height</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intensity</td>
<td>.21</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Size</td>
<td>-.67&lt;sup&gt;a&lt;/sup&gt;</td>
<td>.35&lt;sup&gt;b&lt;/sup&gt;</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Valence</td>
<td>.44&lt;sup&gt;c&lt;/sup&gt;</td>
<td>.28&lt;sup&gt;d&lt;/sup&gt;</td>
<td>-.25&lt;sup&gt;a&lt;/sup&gt;</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Height</td>
<td>.61&lt;sup&gt;d&lt;/sup&gt;</td>
<td>.37&lt;sup&gt;b&lt;/sup&gt;</td>
<td>-.40&lt;sup&gt;a&lt;/sup&gt;</td>
<td>.54&lt;sup&gt;d&lt;/sup&gt;</td>
<td></td>
</tr>
<tr>
<td>Quantity</td>
<td>-.34&lt;sup&gt;b&lt;/sup&gt;</td>
<td>.52&lt;sup&gt;d&lt;/sup&gt;</td>
<td>.75&lt;sup&gt;d&lt;/sup&gt;</td>
<td>.10</td>
<td>-.04</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Non-musicians</th>
<th>Pitch</th>
<th>Intensity</th>
<th>Size</th>
<th>Valence</th>
<th>Height</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intensity</td>
<td>.32&lt;sup&gt;b&lt;/sup&gt;</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Size</td>
<td>-.62&lt;sup&gt;d&lt;/sup&gt;</td>
<td>.27&lt;sup&gt;b&lt;/sup&gt;</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Valence</td>
<td>.40&lt;sup&gt;c&lt;/sup&gt;</td>
<td>.42&lt;sup&gt;c&lt;/sup&gt;</td>
<td>-.23</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Height</td>
<td>.65&lt;sup&gt;d&lt;/sup&gt;</td>
<td>.33&lt;sup&gt;b&lt;/sup&gt;</td>
<td>-.46&lt;sup&gt;b&lt;/sup&gt;</td>
<td>.63&lt;sup&gt;d&lt;/sup&gt;</td>
<td></td>
</tr>
<tr>
<td>Quantity</td>
<td>-.34&lt;sup&gt;b&lt;/sup&gt;</td>
<td>.47&lt;sup&gt;d&lt;/sup&gt;</td>
<td>.73&lt;sup&gt;d&lt;/sup&gt;</td>
<td>.19</td>
<td>-.05</td>
</tr>
</tbody>
</table>

<sup>a</sup> p < .05.<br /><sup>b</sup> p < .01.<br /><sup>c</sup> p < .001.<br /><sup>d</sup> p < .0001.

<table>
<thead>
<tr>
<th>Musicians</th>
<th>Pitch</th>
<th>Intensity</th>
<th>Size</th>
<th>Valence</th>
<th>Height</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intensity</td>
<td>.47&lt;sup&gt;c&lt;/sup&gt;</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Size</td>
<td>-.51&lt;sup&gt;d&lt;/sup&gt;</td>
<td>.46&lt;sup&gt;c&lt;/sup&gt;</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Valence</td>
<td>.12&lt;sup&gt;c&lt;/sup&gt;</td>
<td>.05&lt;sup&gt;c&lt;/sup&gt;</td>
<td>-.25&lt;sup&gt;a&lt;/sup&gt;</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Height</td>
<td>.16&lt;sup&gt;c&lt;/sup&gt;</td>
<td>.30&lt;sup&gt;a&lt;/sup&gt;</td>
<td>-.41&lt;sup&gt;a&lt;/sup&gt;</td>
<td>.22</td>
<td></td>
</tr>
<tr>
<td>Quantity</td>
<td>-.04&lt;sup&gt;c&lt;/sup&gt;</td>
<td>.11&lt;sup&gt;c&lt;/sup&gt;</td>
<td>.64&lt;sup&gt;d&lt;/sup&gt;</td>
<td>.22</td>
<td>.35&lt;sup&gt;b&lt;/sup&gt;</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Non-musicians</th>
<th>Pitch</th>
<th>Intensity</th>
<th>Size</th>
<th>Valence</th>
<th>Height</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intensity</td>
<td>.52&lt;sup&gt;d&lt;/sup&gt;</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Size</td>
<td>-.40&lt;sup&gt;c&lt;/sup&gt;</td>
<td>.40&lt;sup&gt;c&lt;/sup&gt;</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Valence</td>
<td>-.06&lt;sup&gt;c&lt;/sup&gt;</td>
<td>.21&lt;sup&gt;c&lt;/sup&gt;</td>
<td>-.27&lt;sup&gt;a&lt;/sup&gt;</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Height</td>
<td>.31&lt;sup&gt;b&lt;/sup&gt;</td>
<td>.08&lt;sup&gt;c&lt;/sup&gt;</td>
<td>-.28&lt;sup&gt;b&lt;/sup&gt;</td>
<td>.38&lt;sup&gt;b&lt;/sup&gt;</td>
<td></td>
</tr>
<tr>
<td>Quantity</td>
<td>-.16&lt;sup&gt;c&lt;/sup&gt;</td>
<td>.19&lt;sup&gt;c&lt;/sup&gt;</td>
<td>.61&lt;sup&gt;d&lt;/sup&gt;</td>
<td>.30&lt;sup&gt;b&lt;/sup&gt;</td>
<td>.21&lt;sup&gt;c&lt;/sup&gt;</td>
</tr>
</tbody>
</table>

<sup>a</sup> p < .05.<br /><sup>b</sup> p < .01.<br /><sup>c</sup> p < .001.<br /><sup>d</sup> p < .0001.
5.3.2. Appropriateness ratings

The participants also rated the appropriateness of the antonyms for each dimension. The correlations between the ratings for different dimensions are shown in Table 7. Highest correlations with the ratings for pitch height include spatial height, size and intensity. The correlation between pitch height and spatial height is stronger for the non-musician participants than for the musicians. The ratings of valence and pitch correlate negatively, indicating that those antonyms that are strongly associated with valence tend to be weakly associated with pitch and vice versa.

5.3.3. Experiment 1 vs. Experiment 4: pitch metaphors and notions of pitch production

In Experiments 1 and 4 we asked two different questions concerning pitch. While in Experiment 1 participants were requested (task 1) to note which term in each antonym pair is a better metaphor for high (alternatively, low) pitch, in Experiment 4 (Q5, task 1), participants were asked to apply the appropriate terms to objects producing high or low pitch.

To compare the results obtained in this experiment for the judgment of pitch height with the results obtained in Experiment 1, a series of Chi squared analyses were conducted for the 29 antonyms that were common to both experiments (complemented, wherever cells approached 0, by Fisher’s Exact Tests). Following FDR multiple tests corrections, these analyses showed no significant interactions between response distribution and experiment, suggesting similar results in both experiments. Indeed, the correlation between the consensus data in Experiments 1 and 4 (Q5) was very high (r = .87). Notably, responses were strikingly similar even when the qualities in question have nothing to do, physically, with the production of high or low pitch. For instance, large majorities of participants agreed not only that “dark” is an appropriate metaphor for low pitch (Experiment 1), but also assert in Experiment 4 that dark objects tend to produce lower-pitched sounds!

5.4. Discussion

Experiment 4 suggests that qualities associated with pitch production are correlated with qualities related to intensity (higher pitch – higher intensity), spatial height (low pitch – low elevation) and size/mass (negatively: high pitch – small size/mass). The correlation between pitch and spatial height can, however, be largely attributed to the co-correlation with other dimensions, as indicated by a low partial correlation, especially for the musician participant group. Antonyms appropriate for size and intensity were also seen as appropriate antonyms for pitch. Only the non-musicians showed a correlation between the appropriateness of antonyms for pitch height and spatial height. Surprisingly, in this experiment, pitch correlated weakly with valence, and antonyms that were considered appropriate for valence were not considered to be appropriate for pitch.

Experiment 4 points at two intriguing notions. The first (already noted while comparing Experiments 1 and 3) is that pitch may be associated with metaphorical mappings involving magnitude in two contrasting ways. Higher pitch is associated with the “less” pole of the mass/size and quantity dimensions, as it is considered small and lightweight; in contrast, high pitch is related to the “more” pole of other dimensions: spatial height, intensity, and positive valence. We shall further consider this important “conflict of magnitudes” in the general discussion.

Second, pitch metaphors are strongly associated with features ascribed to sound-producing objects. Yet surprisingly, this relationship may be reciprocal. Indeed, some of the metaphors associated with high or low pitch reflect experienced corollaries of pitch height, such as physical size or gender. But the inverse relationship seems to hold as well: we may associate the production of high or low pitch to qualities (like visual brightness) which have nothing to do with it in experienced physical reality, since they commonly serve as metaphors for pitch. Despite repeated experiential evidence to the contrary, we may build a false “folk physics” of sound, based on the metaphors we use to describe it.

6. General discussion

This study supports and broadens the notion, suggested by previous empirical research, that a web of shared connotations, encompassing diverse sensory modes as well as cognitive and social domains, is associated with the concept (and possibly the percept) of auditory pitch. In three experiments, including two conceptual, matching tasks (Experiments 1 and 4), and an adjective rating task using actual musical segments (Experiment 2), different partici-
Experiment 1 used a purely conceptual task, where participants were asked to apply metaphors to the concepts of high and low pitch (in Experiments 1, 2, and 4) was congruent with their original application. This congruence pertained not only to broad characteristics, such as small–large or young–old, but also to seemingly idiosyncratic metaphors like sharp–heavy, grandmother–daughter, and crocodile–those who follow the crocodile. These results suggest that non-Western mappings of pitch, though using culture-specific associations, are also based on more general, cross-cultural connotations, which may stem from basic bodily experiences and interactions with the physical environment. Hence, such mappings may make intuitive sense, even though one might have never been aware of them. For instance, non-Western metaphors like “crocodile” or “grandmother–granddaughter,” which may sound odd to those outside their original cultural milieu, are still readily understood by Western subjects, possibly since they are partially based on the general experiential correlation of pitch height with physical size and mass.

Second, previous empirical studies concerning cross-domain mapping of pitch (see introduction) used very simple stimuli (typically, isolated synthesized pitches), far from the complexity of even the simplest musical segment. These stimuli, nevertheless, inevitably possessed qualities other than pitch (e.g., a specific duration and timbre), which might have interfered with participants’ judgments. Here, we combined and compared two experiments. Experiment 1 used a purely conceptual task, where participants were asked to apply metaphors to the concepts of high and low pitch, rather than to a specific sound presented to them. Experiment 2, in contrast, involved segments from an actual piece of music, which differed primarily in pitch register. Comparison between these two tasks suggests that abstract conceptual mappings of pitch generalize to a considerable extent to mappings of complex musical stimuli.

This finding (which should be supported by further studies, using a wide gamut of musical styles and characteristics) implies that the metaphorical web of pitch connotations investigated in the present study may supply an interesting tool for the analysis and criticism of music. For instance, expressive or programmatic connotations of “high” and “low” musical segments (as revealed by texts, programs, or listeners’ reactions) may be associated with such implicit, but widely-shared, metaphorical web associated with “high” and “low” pitch. Analysis of explicit or implicit cross-domain pitch mapping may also afford a better understanding of verbal discourse concerning music— from theoretical treatises through concert programmes to music-related prose and poetry. Finally, such connotations may help investigating, interpreting and perhaps discovering systematic phenomena concerning pitch register in music. The fact, for instance, that lower pitches are larger, heavier and “older,” may meaningfully relate (both ways) to the different temporal behavior of pitch in different registers (e.g., Tamir & Eitan, 2007).

Though pitch metaphors identified in the abstract are generalizable to actual music, this generalization is restricted by other musical features which may carry their own metaphorical web. Even when other features of the music (like texture, dynamics, or rhythmic complexity) are strictly controlled, these features may interact differently with pitch in different registers, and thus affect its cross-domain mappings. In the present experiment, interactions with features like textural and rhythmic density or perceived loudness probably generated several differences between pitch connotations in the abstract (Experiment 1) and in the context of Beethoven’s music (Experiment 2).

The effect of such interactions on cross-domain mapping of music, and of auditory stimuli in general, has hardly been studied empirically, and would certainly merit further research. Indeed, as Nygaard et al. (2009) have recently reported, different features combine to produce unique semantic “acoustic signatures” in speech prosody. Thus, for instance, “big” (as compared to “small”) is marked by lower pitch, increased loudness, and longer duration, while “happy” (as compared to “sad”) is marked by higher pitch, wider pitch variation, increased loudness, and shorter duration.

An auditory parameter whose effect on the “semantics” of sound (alone and in interaction with pitch) may be of particular interest is loudness. Loudness has been shown to interact perceptually with several non-auditory dimensions, including brightness (already at early infancy; Lewkowicz & Turkewitz, 1980), size (Smith & Sera, 1992) and height (Eitan, Schupak, & Marks, 2008). Moreover, though loudness and pitch themselves interact, such that louder and higher-pitched sound, as well as sounds increasing in loudness and pitch, correspond (Melara & Marks, 1990; Neuhoff & McBeath, 1996; Neuhoff, McBeath, & Wanzie, 1999), “higher” loudness and pitch interact with other dimensions in contrasting ways. Increased loudness, for instance, is “bigger” (Smith & Sera, 1992), while increased (higher) pitch is “smaller.”

6.1. Underlying dimensions

The introduction to this article presented several hypotheses concerning possible underlying dimensions of
cross-domain mappings for pitch. We suggested that, alternatively or complementarily, pitch metaphors may be based on second-order mappings of spatial verticality; on a generalized magnitude representation; or on qualities attributed to pitch-producing objects. Here, we re-examine these hypotheses in light of the four experiments presented above.

6.1.1. Pitch and spatial verticality

We compared cross-domain mappings of pitch with those of spatial verticality in several ways. First, we conducted a between-subjects comparison of the main task in Experiment 1 (in which participants were asked which pole in each of 29 antonyms is a better metaphor for high/low pitch), with that of Experiment 3 (where a similar question was asked with regard to high or low spatial elevation). Second, we asked participants in Experiment 3 to match auditory antonyms, as well as actual sound stimuli differing in pitch register, with high and low elevation. Finally, in the within-subject design of Experiment 4, participants were asked (among other questions) which pole in each of 36 antonyms is a better description of objects producing high/low sounds, and which pole describes spatially high/low objects better.

Two conclusions, seemingly contradictory, are indicated by results of these comparisons. First, participants indeed closely associated high and low pitch with high and low elevation. They consensually matched auditory terms and stimuli with spatial high and low, and highly rated auditory adjectives as metaphors for spatial elevation (Experiment 3); in addition, a strong pairwise correlation between pitch and spatial elevation was found in Experiment 4 (though this relationship was considerably attenuated when partial correlation, which takes into account the other four variables, was analysed). Yet, the associations with pitch and spatial elevation also show strong discrepancies, for example related to physical magnitude: high pitch is small and little, while spatial height is associated with large and much (as found in Experiment 3). In correspondence with this disparity, high pitch is also young and female, while spatial height is associated with old and male. Interestingly, the verticality metaphor is itself not unambiguously related to physical magnitude: although spatial height is associated with larger and more, it is also lighter and thinner.

Another dimension with which the connotations of pitch and spatial height differ is valence. As Lakoff and Johnson (1980b) show, spatial height serves as an important source domain in evaluative metaphors, strongly linked with moral value, rationality, and control. Importantly, these mappings transcend the linguistic domain: they affect, and are affected by, bodily motion and gesture in more than one way (Casasanto, 2008, 2009; see above, pp. 2, 11). In the present study, higher elevation was indeed significantly correlated with positive evaluation (Experiment 4), while the correlation of pitch and evaluation was considerably weaker, and disappeared in partial correlations, suggesting that it may be a by-product of correlations among other dimensions. Comparing Experiments 1 and 3 reveals similar differences: “wise” and “pleasant” are significantly associated with high spatial elevation, but not with high pitch, and “ugly” is significantly related to low elevation, but not to low pitch.

Importantly, then, despite the longstanding, ubiquitous use of the verticality metaphor for pitch in the West (including the analogy of pitch and verticality in Western musical notation), the connotative web of auditory pitch seems partly independent of that metaphor, as it possesses different relationships with physical magnitude and evaluation. Pitch metaphors, then, are not an expression of the up–down orientational metaphor (Lakoff & Johnson, 1980a, 1980b), as often claimed (e.g., Cox, 1999), since the main conceptual metaphors shaping up–down metaphors in other domains, such as “up is more” and “up is good” do not generally apply to pitch.

6.1.2. Pitch and magnitude representation: which end is more?

The contrasting relationships of size and spatial height with pitch point to a more general issue. Pitch and other polar dimensions, as represented in results of this and previous studies, correspond in magnitude: increase or decrease in pitch are associated with increase or decrease in other dimensions, such as spatial height, intensity or brightness. Such magnitude analogies are of course not unique (as psychophysical studies since Weber and Fechner indicate) and many (e.g., Cohen Kadosh et al., 2008; Walsh, 2003) have suggested that a shared magnitude representation relates the processing of diverse dimensions and modalities.

However, if pitch metaphors are based on such overall magnitude representation, which end is “more”? Our results (in particular the correlations in Experiment 4) suggest that the percept of pitch involves two contrasting magnitude representations. On the one hand, as pitch “rises” its metaphorical height, intensity, and visual lightness increase; on the other hand, however, its metaphorical mass, size, and quantity decrease. Walker and Smith’s Stroop-like experiments (1984, 1985, 1986) present a similar dichotomy. Higher pitches, then, are both “more” and “less” than lower pitches.

Notably, this “contradiction” is not unique to pitch, as magnitudes in other dimensions are inversely related in a similar way. Walker and Smith (1985) found that smaller haptic size, like higher auditory pitch, is (among other qualities) fast, tense, and bright. Cohen Kadosh, Henik, and Walsh (2007) show that in synesthesia increase in brightness is associated with decrease in numerical magnitude; similarly, Smith and Sera (1992) found that 2 years old children associate increased brightness with smaller objects. Current models of shared magnitude representation may need to take into account and explain such inverse magnitude relationships. Of particular interest is the finding that magnitude relationships in cross-domain mappings are not transitive. For instance (see Table 6) higher pitch is positively correlated with higher intensity, and higher intensity is positively correlated with larger size, but larger size is negatively correlated with higher pitch. Such intransitive relationships indicate that the cross-domain mappings of pitch are underlined by several, sometimes conflicting conceptual metaphors (e.g., “pitch level is intensity level,” in which increased pitch correlates
with increased intensity, vs. “pitch level is size,” in which increased pitch correlates with decreased size).

6.1.3. Pitch metaphors and pitch production

Are pitch metaphors derived, directly or indirectly, from our experience with pitch-producing objects, particularly animate objects? While some of the qualities associated with pitch height (e.g., size) normally correspond to the qualities of objects – from animals to musical instruments – producing high or low pitch, others (such as visual lightness) clearly do not. An object would not produce a lower pitch if painted black, rather than white. Lightness, however, has been shown to strongly associate perceptually or cognitively with pitch in this and several other studies, involving adults (Hubbard, 1996; Marks, 1987; Walker & Smith, 1984, 1986) and children as young as 2.5 years, for whom lightness was the strongest and most consistent mapping of pitch (Marks et al., 1987; Mondloch & Maurer, 2004).

Presumably, then, responses to a task involving application of metaphors to pitch (Experiment 1) would differ, when concerning qualities such as visual lightness or seasons of the year (summer vs. winter), from those of a task indicating qualities actually associated with pitch-producing objects (Experiment 4). This has not been the case. Rather, Chi² analyses comparing matching of antonym pairs with pitch in Experiments 1 and 4 revealed no significant interactions between experiment and response distributions. Correspondingly, the response distributions correlated highly ($r = .87$). Participants indicated not only that “dark” and “winter” are appropriate metaphors for low pitch (Experiment 1), but also that dark objects tend to produce lower pitch, and that lower pitch is produced in wintertime. It seems, then, that we do not only derive pitch metaphors (such as small–large) from pitch production or pitch-producing objects (Experiment 4), but may base our “folk physics” of pitch production on cross-modal mappings of pitch, even when these have no base in physical reality.

In the complexity of its relationships with non-auditory domains, auditory pitch demonstrates how a basic percept may intricately connotes to diverse, seemingly remote realms of experience. These cross-domain mappings are often shared consensually while not explicitly expressed in the vocabulary. Moreover, while often strongly bonded to the specific time and culture that generated them, the underlying dimensions and structure of pitch metaphors may transcend specific cultural contexts. Primal modal aspects of the meaning of sound, reciprocally shaped by and shaping our basic experience with sounding objects, may thus underline even a most complex auditory artefact, music.

However, the significance of exploring the cross-domain connotations of pitch, as well as those of other auditory dimensions, transcends the study of musical “meaning.” The notion that thinking, including the creation and use of abstract concepts, is rooted in embodied perception, has received considerable theoretical and empirical support in recent decades (e.g., Barsalou, 1999; Boroditsky & Prinz, 2008; Lakoff & Johnson, 1980a, 1980b). Revealing how sound relates, on the one hand, to embodied experience involving other sense modalities, and, on the other hand, to concepts involving emotion, evaluation, and even social structure, may increase, but perhaps also problematize, our understanding of how perception and thought interact.

Acknowledgments

We are thankful to Richard Ashley, Candace Brower, Robert Gjerdingen, and Yeshayahu Shen for useful comments and practical assistance. Thanks are also due to David Steinberg and Ilana Galenter for assistance in statistical analysis, and to Mordechai Adler and Golan Gur for their help in data collection. This study was supported by Israel Science Foundation Grant No. 800/02-27.0 to the 1st author and an Israel–USA Binaional Science Foundation (BSF) Grant No. 2005-524 to the 1st author and Lawrence Marks. A partial version of the study was presented at the 9th International Congress on Music Perception and Cognition (ICMPC9), Bologna, Italy, August 22–26, 2006.

References


