Mental Concerts: Musical Imagery and Auditory Cortex

Author links open overlay panel
Robert J. Zatorre [1], Andrea R. Halpern [2]
Show more

Add to Mendeley
Share
Cite

https://doi.org/10.1016/j.neuron.2005.06.013
Get rights and content

Most people intuitively understand what it means to "hear a tune in your head." Converging evidence now indicates that auditory cortical areas can be recruited even in the absence of sound and that this corresponds to the phenomenological experience of imagining music. We discuss these findings as well as some methodological challenges. We also consider the role of core versus belt areas in musical imagery, the relation between auditory and motor systems during imagery of music performance, and practical implications of this research.

Previous article
Next article

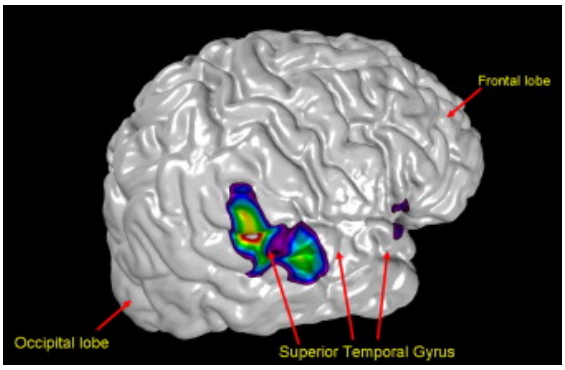
Main Text

Cognitive neuroscientists are faced with a seemingly daunting task: understanding how the brain enables us to experience our rich inner world of thoughts, feelings, and images. The subjectivity involved in these internal processes provides a particular challenge because scientific methods require one to measure verifiable, observable events. One domain in which this problem has played out is mental imagery. Although behaviorists in their heyday insisted that imagery was off limits because of its obscure, subjective nature, clever cognitivists demonstrated early on that reliable behavioral measures could be obtained that served as indices of what was going on inside the mind. Shepard's classic demonstration of mental rotation (Shepard and Metzler, 1971) serves as an excellent example of how an overt measurement (response time to judge the orientation of a letter) can provide evidence of a covert mental process. In other words, one infers the existence of a process based on observing some effect caused by that process; in this respect, cognitive approaches are not so different from physics or other sciences in which the objects of study (neutrinos, black holes, or whatever) are simply not accessible.

Recent Advances in the Study of Musical Imagery

Imagery is not exclusively visual, as anyone can attest to who has ever been annoyed by some advertising jingle playing relentlessly in his or her mind. On a more exalted level, composers such as Beethoven or Smetana, who became deaf later in their lives, nonetheless were able to compose magnificent music, presumably because they were able to conjure up musical images solely internally. Many researchers have concentrated on understanding musical imagery in particular partly because of the ubiquity and vividness of imagined music. So what enables us to produce these "mental concerts"? What are the psychological and neural mechanisms associated with these processes?

A handful of studies have now been carried out on this topic using a variety of techniques, including, magneto-encephalography (Schürmann et al., 2002), positron emission tomography (Halpern and Zatorre, 1999, Zatorre et al., 1996), and functional MRI (Halpern et al., 2004, Kraemer et al., 2005, Yoo et al., 2001), as well as behavioral lesion measures (Zatorre and Halpern, 1993), which provide better evidence of causality than do functional measures. These diverse studies converge on one principal finding: that neural activity in auditory cortex can occur in the absence of sound (Figure 1) and that this activity likely mediates the phenomenological experience of imagining music. Beyond this basic understanding, however, much remains to be understood, including the relative contributions of primary versus secondary auditory regions in each hemisphere, the participation of the frontal cortex to the imagery process, and the role of musical training in development of musical imagery. Before discussing these substantive questions, however, we turn our attention briefly to methodological issues.



Download : Download high-res image (199KB)
Download : Download full-size image

Figure 1. Lateral View of Right Cerebral Hemisphere Illustrating Area of Hemodynamic Increase, in Color, during an Auditory Imagery Task

Although the task is performed in silence, activation is observed within auditory cortex in the posterior aspect of the superior temporal gyrus. Data reanalyzed from Halpern et al., 2004.

Methodological Problems and Solutions

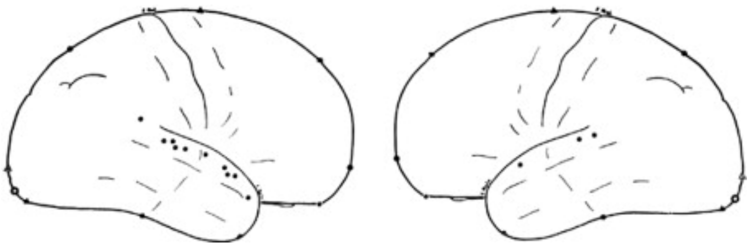
The problem of measuring internal phenomena might appear to have finally found a solution with functional imaging techniques, since one can observe the underlying neural activity more directly, rather than inferring its presence. Yet, we are still left with the conceptual problem of knowing what is being measured. Thus, merely placing subjects in a scanner and asking them to imagine some music, for instance, simply will not do, because one will have no evidence that the desired mental activity was actually taking place. Neural activity can still be measured under these circumstances, but it may well be related to other processes than the one intended. One good solution to this problem involves behavioral indices, such that an overt response is measured that either depends on or correlates with the imagined event. For instance, if we ask people to imagine the first four notes of Beethoven's Fifth Symphony and they correctly and consistently judge that the fourth note is lower than the third, then we have objective evidence that an internal representation containing pitch information has been accessed (Halpern, 1988). In the context of neuroimaging, such tasks have the disadvantage that they carry a lot of cognitive overhead in the form of attentional, working memory, and response demands; but these can be accounted for with appropriate control conditions. Simpler tasks, such as imagining the continuation of a known musical selection (Kraemer et al., 2005) can also be useful. But in such tasks there is less control over the success with which imagery may be achieved at any given moment, absent a behavioral correlate, such as the time taken for the continuation, matched against the length of the excerpt. Using fMRI poses special problems because of the loud acoustical artifact produced by echo-planar imaging, which itself results in a large auditory cortical response. The interactions between this response and the one related to the formation of auditory imagery makes interpretation difficult, particularly when the hemodynamic response functions between imagined and perceived events overlap (Kraemer et al., 2005). This problem can be mitigated, however, by using sparse sampling or other noise abatement strategies.

What Role Does Auditory Cortex Play in Imagery?

Despite these technical difficulties, most imagery studies have indeed succeeded in demonstrating that the auditory cortex responds even in the absence of sound and that this response tends to co-occur with subjective reports of imagining music. But does the primary, or core auditory cortex, participate in musical imagery? A similar question exists in the field of visual imagery, but there is now substantial evidence that primary visual cortex can be recruited by certain tasks (Kosslyn and Thompson, 2003). The literature in musical imagery to date is still uncertain on this point. Most prior studies do agree that activation in secondary, or belt, auditory cortex is reliably found (Figure 1). Although some authors have reported activation in primary cortex, the precise location of core areas can be difficult to determine because of the intersubject variability of these structures; furthermore, because of partial volume effects, what may appear to be activity in primary regions may actually represent spillover from adjacent nonprimary zones. The most critical variable, however, is likely to be the task demands, as they have proven to be in the visual domain. It is premature to anticipate what auditory imagery tasks might reliably elicit primary activation, given the very small number of studies carried out so far and the small subset of those that have even sought to verify the precise location of the activation.

An additional point that many of these studies address concerns the lateralization of the response; once again, task demands and also the nature of the stimuli to be imagined likely play a role. For instance, bilateral activation has been observed when familiar songs with lyrics are used, most likely because there is imagery of both the sung text and the musical component (Zatorre et al., 1996). But when instrumental music is used (Halpern et al., 2004), the pattern tends to shift toward activation in the right auditory cortex, in accord with the important role of these structures in processing pitch information (Zatorre et al., 2002). The recent study by Kraemer et al. (2005) did show left auditory cortex activation even with nonverbal materials, but the degree of activity on the right, if any, was not reported, leaving this question still open.

Assuming we can agree that auditory cortical activity underlies the experience of imagery, the question still remains, how does the auditory cortex become active in the first place? The most likely explanation is that top-down mechanisms are involved in reactivating neural traces that are somehow encoded in sensory cortex. Long ago, Penfield observed that electrical stimulation of the exposed surface of sensory cortical areas (Figure 2) could result in the patient reporting illusory visual or auditory percepts (Penfield and Perot, 1963). The artificial electrical input from an electrode results in a hallucinatory rather than an imagery experience, but presumably under normal circumstances there is a signal coming from elsewhere that accesses the sensory information in auditory cortex. It is most likely that interactions between frontal cortical areas and auditory cortex are the way that imagery is instantiated. There is a tight anatomical connectivity between these regions, and most studies that report whole-brain data involving the generation of an auditory image find frontal cortex to be an important component (Halpern and Zatorre, 1999). Thus, when one wants to conjure up a song in one's mind, frontal-based retrieval mechanisms might be called upon; at the same time, feedback signals from auditory cortex could be important in distinguishing between imagery and a real sound coming from the environment. Indeed, Griffiths (2000) proposed that a breakdown in this system might be responsible for the musical hallucinations that he observed in people with acquired deafness. It is notable that this study found no evidence for primary auditory cortex activation, making suspect any argument linking primary cortex activation with stronger phenomenology of imagined sounds.



Download : Download high-res image (85KB)
Download : Download full-size image

Figure 2. Regions of the Exposed Cortical Surface, Marked with Dots, which Resulted in a Hallucinatory Experience of Hearing Music Upon Electrical Stimulation

Modified from Penfield and Perot, 1963.

Auditory versus Motor Imagery

Motor imagery is the imagination of the kinesthetics involved in actual movement and has been examined for both simple tapping sequences and complex musical routines. One methodological challenge in examining brain activations in motor imagery is to insure that no actual movements have occurred, which can be accomplished via EMG monitoring. Motor imagery for nonauditory-associated movements sometimes results in activation of M1 and often activates secondary motor areas, such as SMA (Naito et al., 2002). Thus it should not be surprising that musicians can evoke motor imagery for their instrument during imagined playing. For instance, Langheim et al. (2002) asked string players to play or imagine playing a familiar piece; the times taken to play and imagine the pieces were highly correlated. These authors found a number of areas to be active in frontal lobes, cerebellum, parietal lobe, and SMA, but not M1 during imagined playing compared to rest.

In many musical situations, sound is associated with movement. Instrumentalists make extensive arm, finger, and sometimes foot movements in the course of producing their instrument's voice. Singers use complex movements of the vocal apparatus to produce songs, especially if they are putting words to those songs. Given the behavioral and neural evidence for people being able to imagine musical movements, is there evidence that auditory and motor imagery may be integrated in the brain? Hickok et al. (2003) found that area Spt (parietal-temporal boundary) responded to both imagined auditory (both speech and music) and covertly produced sequences in a similar fashion. In perhaps a stronger test of this integration, Haueisen and Knösche (2001) found that pianists showed activation in primary motor regions corresponding to the finger that would have produced a given note, even when they were merely listening to pieces they knew how to perform. Conversely, Haslinger et al. (2005) observed activation in several auditory areas when musicians watched a silent video of someone fingering piano keys. Thus, despite a rather different circuitry, imagery of related musical sounds and movement can be integrated. This corresponds to reports from musicians that they can "hear" their instrument during mental practice.

Cross-Modal Interactions

Processing in one sensory modality can affect processing in another, either by increasing or suppressing activity; similar interactions also appear to occur if one or both tasks are based not on perceptual, but on imagined information. Langheim et al. (2002) found that imagining musical performances suppressed activity in the auditory regions, although they suggested that it may have been related to suppression of scanner noise. Halpern et al. (2004) also found that a visual imagery task suppressed activity in right secondary auditory cortex (which was active in imagery for musical timbre) to levels below that seen with a silent baseline. As noise was not a factor given the sparse sampling technique used, it seems that cross-modal interactions may operate similarly in auditory imagery as they do in the processing of actual sound.

Implications

Musical imagery is important to musicians, so an understanding of its neural basis may help us understand aspects of expertise as well as provide some useful information for music educators. For instance, brass, wind, string players, and singers imagine the pitch of an upcoming entrance to facilitate tuning. Conductors and arrangers who study scores in silence also must imagine pitches, as well as timbre, rhythm, and other musical attributes. Highben and Palmer (2004) asked pianists to learn an unfamiliar piece, under normal conditions or when they could not hear their own playing. Players who tested high on an aural skills battery were least disturbed in learning the piece without auditory feedback, suggesting that their auditory imagery skills were adding in the necessary auditory experience to facilitate learning. Humphreys (1986) reported that training in auditory imagery improved harmony skills in children.

The research described above can also help illuminate how musicians use mental practice. This skill involves imagery in several modalities: visual (pianists "see" their hands on the keyboard), motor/kinesthetic (they "feel" the keyboard and finger motions), as well as auditory. Experimental evidence bearing on the neural processes involved is still quite limited, but Pascual-Leone (2003) has demonstrated that mental practice improves performance, albeit not to the level of real practice. However, changes over time in the size of the cortical representation of the motor cortex were similar for real and imagined practice. Given the existence of cross-modal interactions, we may eventually be in a better position to explain when and how these imagined experiences will actually benefit musicians and thus be able to optimize practice regimes for individuals as well as add to the literature on neuroplasticity in response to expert training.

We have attempted here to argue that well-considered behavioral methods combined with convergent neuroimaging and other techniques can successfully externalize the particularly covert process of musical auditory imagery. This research allows us to gain insight into one of the more inaccessible aspects of cognition, and thereby provides us with valuable information concerning the neural underpinnings of abstract mental processes. Clinical or educational applications pertaining to these highest levels of cognitive function will emerge only to the extent that we can rigorously link brain mechanisms to mental processes; we would argue that the future of cognitive neuroscience will depend on expanding just this sort of knowledge.

(e.g. in Word or Tex form) to their personal website or institutional repository. Authors requiring further information regarding Elsevier's archiving and manuscript policies are encouraged to visit: http://www.elsevier.com/copyright Author's personal copy 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 bDepartment of Music, University of Sheffield, United Kingdom article info Article history: Received 14 February 2009 Revised 3 September 2009 Accepted 14 October 2009 Keywords: Cross-domain mapping Metaphor Auditory pitch Music cognition 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 latently 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. 2009 Elsevier B.V. All rights reserved. 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 0010-0277/$ - see front matter 2009 Elsevier B.V. All rights reserved. doi:10.1016/j.cognition.2009.10.013 * Corresponding author. Tel.: +972 548100998. E-mail address: ZEITAN@POST.TAU.AC.IL (Z. Eitan). Cognition 114 (2010) 405–422 Contents lists available at ScienceDirect Cognition journal homepage: www.elsevier.com/locate/COGNIT Author's personal copy 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, Umilta, & 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 retraced 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). 406 Z. Eitan, R. Timmers / Cognition 114 (2010) 405–422 Author's personal copy 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 & Granot, 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   reached for alert–sleepy, sharp–heavy, thin–thick, young– old, sharp–blunt and fast–slow. 2.3.1.1. 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. Those antonyms that showed a significant interaction ($p < .03$) are marked in Table 1 (As column). For three antonyms, choices for high and low pitch were asymmetrical: the proportion of participants who chose a term as appropriate for high pitch was significantly different from the proportion of participants who chose its opposite as appropriate for low. For two antonyms (pretty–ugly and loud–soft) consensus was higher for high pitch, while for another (empty–full) consensus was higher for low pitch. In other words, high pitch is loud and pretty, but low pitch is not soft and ugly, while low pitch is full, but high pitch is not empty. These asymmetries form intuitive exceptions to the general symmetry of the consensus for the other antonyms. 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 intermediately 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 antonym 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 variation 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 appropriateness 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). Nevertheless, a few ratings showed a significant asymmetry, as assessed by a paired samples T-test (as in Table 1): the antonyms alert–sleepy and tense–relaxed were rated significantly 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.). Consensus Appropriateness H L As. H L As. Active–passive .87d .30b 2.8 2.6 Alert–sleepy 1.0d .10d 3.4 3.1 Beginning–end .66a .23d 2.7 2.6 Cold–hot .72c .28c 3 3.1 Empty–full .62 .20d 2.3 3.1 Far–near .58 .37 2.7 2.7 Fast–slow .95d .05d 3.3 3.3 Feminine–masculine .90d .07d 3.1 3.2 Fool–wise .51 .28c 2.2 2.3 Grdaugh–Grdma .93d .08d 2.4 2.4 Happy–sad .95d .03d 3.1 3 Hard–soft .57 .47 2.6 2.8 Light–dark .95d .05d 3.3 3.3 Light–heavy .95d .11d 3.3 3.6 Little–much .57 .36a 2.4 2.7 Loud–soft .76c .51 2.8 2.6 Pleasant–unpleasant .53 .63a 2.7 2.5 Pretty–ugly .86d .42 2.1 2.2 Right–left .75c .32b 2.1 2.2 Rough–smooth .85d .18d 3.5 3.3 Sharp–blunt .97d .08d 3 3 Sharp–heavy .98d 0 d 2.9 3.1 Small–large .90d .15d 2.7 2.5 Sparse–dense .72c .23d 2.9 2.9 Strong–weak .64a .54 2.8 2.9 Summer–winter .80d .18d 2.7 2.7 Tense–relaxed .90d .17d 3.2 2.9 Thin–thick .98d .03d 3.8 3.9 Young–old .97d .10d 3.1 2.9 Proportion of participants choosing high or low is significantly different from 50%. a p < .03. b p < .01. c p < .001. d p < .0001. Z. Eitan, R. Timmers / Cognition 114 (2010) 405–422 409 Author's personal copy 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 crossdomain 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. 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 averagemore 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 CZS5729122). 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 metaphoric 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.   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; $p < 0.05$).1 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 highpitched 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 Table 2 Overview of Experiment 2 results. Listed are the ratings for the highpitched (H) and low-pitched (L) music, the difference between these ratings (D), 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). Antonym H L D p Diff. Exp. 1 Active–passive 3.4 3 0.4 Alert–sleepy 3.8 2.6 1.2 d Beginning–end 2.8 1.6 1.2 d Cold–hot 2.1 1.6 0.5 Empty–full 2.2 1.2 1 d Far–near 3 3 0 Fast–slow 3.7 2.4 1.3 d Feminine–masculine 3.1 1.1 2 d Fool–wise 1.6 1.4 0.2 Grdaugh–Grdma 3.2 1.1 2.1 d Happy–sad 3.7 2.4 1.3 d Hard–soft 2.1 3.1 1 d Light–dark 3.4 0.8 2.6 d Light–heavy 3.5 1 2.5 d Little–much 2.4 1.7 0.7 a Loud–soft 2.3 3.1 0.8 d Pleasant–unpleasant 4 3.3 0.7 b Pretty–ugly 4.2 3.3 0.9 d Right–left 3.4 2.9 0.5 Rough–smooth 1.9 3.5 1.6 d Sharp–blunt 3.8 2 1.8 d Sharp–heavy 3.7 1.9 1.8 d Small–large 3.1 1.3 1.8 d Sparse–dense 2.6 1.3 1.3 d Strong–weak 2.5 3.4 0.9 b Summer–winter 3.9 2 1.9 d Tense–relaxed 2.8 3.4 0.6 Thin–thick 3.2 0.8 2.4 d Young–old 3.2 1.1 2.1 d Cause–result 2 1.9 0.1 Clean–dirty 4.4 2.9 1.5 d Follow–crocodile 2.3 1.4 0.9 a Good–bad 4 3.1 0.9 d Sane–crazy 2.6 2.4 0.2 Sweet–sour 4.1 2.6 1.5 d a $p < .05$. b $p < .01$. c $p < .001$. d $p < .001$. 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 (Chi2 = 40, $p < .0001$; Chi2 = 30, $p < .0001$). These were followed by the antonyms hard–soft, strong–weak, alert–sleepy and pleasant– unpleasant (Chi2 = 15, $p < .0001$; Chi2 = 14, $p < .0001$; Chi2 = 11, $p < .01$; Chi2 = 11, $p < .01$). For hard–soft, strong–weak, and pleasant–unpleasant the association between pitch and the poles of the antonyms was indecisive in Experiment 1, but decisive in Experiment 2. While for alert–sleepy and tense–relaxed, 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, Succi, & 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 Table 3 Summary of results of the factor analysis on differences in antonym rating between the high and low music conditions (Experiment 2). To characterize the factors, antonyms demonstrating a strong correlation with the factors are listed ($r > .50$, $p < .0001$). The first term of each antonym is associated with the highpitched music. F1 F2 F3 F4 Antonym r Antonym r Antonym r Antonym r Thin–thick .77 Weak–strong .79 Pleasant–unpleasant .80 Alert–sleepy .79 Sharp–blunt .72 Soft–loud .77 Good–bad .74 Fast–slow .75 Feminine–masculine .60 Little–much .68 Sweet–sour .72 Grdaugh–Grdma .55 Sharp–heavy .68 Pretty–ugly .71 Follow–crocodile .55 Light–dark .59 Hard–soft .65 Light–heavy .58 Clean–dirty .59 Smooth–rough .58 Light–dark .52 Small–large .58 Young–old .55 412 Z. Eitan, R. Timmers / Cognition 114 (2010) 405–422 Author's personal copy 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.2 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 experientially 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, Hammeal, & Bornstein, 1987; Mondloch & Maurer, 2004).3 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 crossmodal, 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 Verticality, 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 2 A recently completed experiment similar in design to Experiment 2 (Eitan & Adler, 2008), using segments from Bach's Goldberg Variations, largely corroborates the present results. 3 Brightness and lightness are not identical dimensions. However, in Hebrew, used in this experiment, the same word (''bahir'') is sometimes used to denote both terms.   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 by 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.4 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 rated 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. 4 Note that since in Hebrew the same word (Gavoha) is used for both ‘‘tall" and ‘‘high," participants were specifically instructed to apply metaphors to the dichotomy high–low, as distinguished from tall–small. 414 Z. Eitan, R. Timmers / Cognition 114 (2010) 405–422 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 verticalitymetaphors (and, asmentioned, 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 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 Exps. 1 (pitch) and 3 (spatial elevation) are compared. Consensus Appropriateness Exp. 3 H Exp. 1 H Exp. 3 L Exp. 1 L Exp. 3 H Exp. 1 H Exp. 3 L Exp. 1 L Active–passive .86 .87 .14 .30 2.67 2.8 2.41 2.6 Alert–sleepy .91 1.0 .14 .10 2.74 3.4b 2.82 3.1 Beginning–end .24 .66a .65 .23a 2.81 2.7 2.82 2.6 Cold–hot .51 .72 .39 .28 2.97 3 2.59 3.1 Empty–full .37 .62a .41 .20a 2.81 2.3 2.67 3.1 Far–near .88 .58 .19 .37a 3.14 2.7 2.91 2.7 Fast–slow .91 .95 .1 .05 2.85 3.3 2.75 3.3b Feminine–masculine .35 .90a .61 .07a 2.57 3.1 2.21 3.2b Fool–wise .11 .51a .85 .28a 2.36 2.2 2.28 2.3 Grdaugh–Grdma .39 .93a .56 .08a 2.43 2.4b 2.37 2.4b Happy–sad .84 .95 .1 .03 3.16 3.1 2.82 3 Hard–soft .6 .57 .48 .47 2.31 2.6 2.33 2.8 Light–dark .78 .95a .21 .05a 2.97 3.3 2.66 3.3b Light–heavy .85 .95 .17 .11a 3.14 3.3 3.31 3.6 Little–much .25 .57a .77 .36a 3.02 2.4b 3.05 2.7 Loud–soft .84 .76a .21 .51a 3.4 2.8 3.34 2.6b Pleasant–unpleasant .88 .53a .27 .63a 2.26 2.7 2.16 2.5 Pretty–ugly 1 .86 .13 .42a 2.1 2.1 2.1 2.2 Right–left .7 .75 .29 .32 1.66 2.1 1.53 2.2b Rough–smooth .83 .85 .21 .18 1.9 3.5b 1.98 3.3b Sharp–blunt .89 .97 .09 .08 2.71 3.0b 2.3 3.0b Sharp–heavy .86 .98a .1 0 2.54 2.9 2.4 3.1 Small–large .18 .90a .77 .15a 3.78 2.7b 3.41 2.5 Sparse–dense .71 .72 .28 .23 3.07 2.9 2.91 2.9 Strong–weak .84 .64a .21 .54a 3.11 2.8 2.9 2.9 Summer–winter .65 .80 .24 .18 2.5 2.7 2.21 2.7 Tense–relaxed .86 .90 .11 .17 3.07 3.2b 2.88 2.9 Thin–thick .72 .98a .21 .03a 2.81 3.8 2.97 3.9b Young–old .65 .97a .29 .10a 2.78 3.1 2.69 2.9 Scream–sigh .98 .02 3.14 3.02 Shout–whisper .88 .12 3.4 3.05 Sneeze–cough .74 .26 2.21 2.05 Soprano–bass .98 .02 4.07 3.9 Motif1 .95 .05 3.91 3.97 Motif2 .96 .08 3.97 3.85 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.   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.5) ‘‘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)? 416 Z. Eitan, R. Timmers / Cognition 114 (2010) 405–422 Author's personal copy II. To what degree is each pair associated with spatial height? Q.5 – Pitch height I. Which object would tend to produce a higherpitched 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 nonmusicians 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 5 Correlations between consensuses in the choice of antonyms with respect to the six dimensions of Experiment 4. High and low conditions are both included (N = 70). Results are split for musicians (top) and non-musicians (bottom). Pitch Intensity Size Valence Height Musicians Intensity .21 Size .67d .35b Valence .44c .28a .25a Height .61d .37b .40c .54d Quantity .34b .52d .75d .10 .04 Non-musicians Intensity .32b Size .62d .27b Valence .40c .42c .23 Height .65d .33b .46c .63d Quantity .34b .47d .73d .19 .05 a p < .05. b p < .01. c p < .001. d p < .0001. Table 6 Partial correlations between consensuses in antonym-choice with respect to the six dimensions of Experiment 4. High and low conditions are both included (N = 70). Results are split for musicians (top) and non-musicians (bottom). Pitch Intensity Size Valence Height Musicians Intensity .47c Size .51d .46c Valence .12 .05 .25a Height .16 .30a .41c .22 Quantity .04 .11 .64d .22 .35b Non-musicians Intensity .52d Size .40c .40c Valence .06 .21 .27a Height .31b .08 .28a .38b Quantity .16 .19 .61d .30b .21 a p < .05. b p < .01. c p < .001. d p < .0001. Z. Eitan, R. Timmers / Cognition 114 (2010) 405–422 417 Author's personal copy 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 (Q.5, 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 Chi2 analyses were conducted for the 29 antonyms that were common to both experiments (complemented, wherever cells approached 0, by Fisher's Exact Tests).5 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 (Q.5) 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 particiTable 7 Correlations between the appropriateness ratings for the six dimensions of Experiment 4. High and low conditions are both included (N = 70). Results are split for musicians (top) and non-musicians (bottom). Pitch Intensity Size Valence Height Musicians Intensity .37b Size .38b .39c Valence .26a .12 .32b Height .24a .46c .60d .03 Quantity .14 .32c .76d .22 .52d Non-musicians Intensity .46d Size .46c .36b Valence .19 .02 .29a Height .47d .42c .64d .14 Quantity .21 .44c .71d .10 .48d a p < .05. b p < .01. c p < .001. d p < .0001. 5 To enable comparisons between Experiments 1 and 4, we converted results using the following procedures. In Experiment 1, where participants repeated the same tasks for both high and low pitch, inconsistent responses (where the same metaphor was considered more appropriate than its antonym for both high and low pitches) were excluded from the data, and the remaining data for high pitch was considered. In Experiment 4 (where different participants rated high and low pitch), low pitch responses were reversed (e.g., response choosing ‘‘large” for low pitch was converted into ‘‘small” for high pitch). 418 Z. Eitan, R. Timmers / Cognition 114 (2010) 405–422 Author's personal copy pant groups (musically trained and untrained subjects), similarly, and often almost unanimously, associated pitch with non-auditory domains. These included physical dimensions such as size, mass, visual lightness and texture, the biological and social domains of age and gender, evaluative measures such as happiness and beauty, and facets of activity and force. Thus, higher pitches are interpreted, among other things, as smaller, visually lighter, more lightweight, thinner, sharper and smoother than lower pitches; they are also happier, more alert, younger and female. The high overall agreement among participants for these and other pitch metaphors (including metaphors originating in non-Western cultures) suggests that diverse and strong cross-domain mappings for pitch, not employed in participants' lexicon, exist latently besides the common verticality (‘‘high” and ‘‘low”) metaphor. While many of these mappings have been previously reported (e.g., Walker & Smith, 1984, 1986), the present study examined several underexplored issues pertinent to the understanding of cross-domain mapping of auditory pitch. First, all experiments included pitchmetaphors of non-Western origins, of which participants had no prior knowledge. Western participants' matching of these metaphors with high or 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, suchmappingsmaymake 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 crossdomain 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 Z. Eitan, R. Timmers / Cognition 114 (2010) 405–422 419 Author's personal copy 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 analised). 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 correlates, 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 connote 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. Primeval crossmodal 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 Binational 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 Arom, S., & Voisin, F. (1998). Theory and technology in African music. In The Garland Encyclopedia of world music. I: Africa (pp. 254–270). New York: Garland. Ashley, R. (2004). Musical pitch space across modalities: Spatial and other mappings through language and culture. In S. Lipscomb, R. Ashley, R. Gjerdingen, & P. Webster (Eds.), Proceedings of international conference on music perception and cognition. Adelaide: Causal Productions. Barsalou, L. W. (1999). Perceptual symbol systems. Behavioral & Brain Sciences, 22, 577–660. Ben-Artzi, E., & Marks, L. E. (1995). Visual-auditory interaction in speeded classification: Role of stimulus difference. Perception & Psychophysics, 57(8), 1151–1162. Benjamini, Y., & Hochberg, Y. (1995). Controlling the false discovery rate: A practical and powerful approach to multiple testing. Journal of the Royal Statistical Society, Series B, 57, 289–300. Berlin, B. (1994). Evidence for pervasive synesthetic sound symbolism in ethnozoological nomenclature. In L. Hinton, J. Nichols, & J. J. Ohala (Eds.), Sound symbolism (pp. 76–93). Cambridge, England: Cambridge University Press. Berliner, P. (1978). The soul of Mbira: Music and traditions of the Shona people of Zimbabwe. Berkeley: University of California Press. Bernstein, I. H., & Edelstein, B. A. (1971). Effects of some variations in auditory input upon visual choice reaction time. Journal of Experimental Psychology, 87, 241–246. Boroditsky, L., & Prinz, J. (2008). What thoughts are made of? In G. R. Semin & E. R. Smith (Eds.), Embodied grounding: Social, cognitive, affective, and neuroscientific approaches. New York: Cambridge University Press. Bregman, A. S., & Steiger, H. (1980). Auditory streaming and vertical localization: Interdependence of ‘‘what” and ‘‘where” decisions in audition. Perception and Psychophysics, 28, 539–546. Cabrera, D., Ferguson, S., Tilley, S., & Morimoto, M. (2005). Recent studies on the effect of signal frequency on auditory vertical localization. In Proceedings of ICAD 05 – Eleventh meeting of the international conference on auditory display, Limerick, Ireland, July 6–9, 2005. Casasanto, D. (2008). Conceptual affiliates of metaphorical gestures. Proceedings of the International Conference on Language, Communication, & Cognition. Brighton, UK. Casasanto, D. (2009). Embodiment of abstract concepts: Good and bad in right- and left-handers. Journal of Experimental Psychology: General, 138, 351–367. Casasanto, D., & Dijkstra, K. (submitted for publication). Motor action and emotional memory. Cognition. Casasanto, D., Phillips, W., & Boroditsky, L. (2003). Do we think about music in terms of space: Metaphoric representation of musical pitch. In Proceedings of 25th annual conference of the cognitive science society, Boston, MA. Chuenwattanapranithi, S., Xu, Y., Thipakorn, B., & Maneewongvatana, S. (2008). Encoding emotions in speech with the size code – A perceptual investigation. Phonetica, 65, 210–230. Z. Eitan, R. Timmers / Cognition 114 (2010) 405–422 421 Author's personal copy Cohen Kadosh, R., Henik, A., & Walsh, W. (2007). Small is bright and big is dark in synaesthesia. Current Biology, 17, R834–R835. Cohen Kadosh, R., Lammertyn, J., & Izard, V. (2008). Are numbers special? An overview of chronometric, neuroimaging, developmental and comparative studies of magnitude representation. Progress in Neurobiology, 84, 132–147. Cox, A. W. (1999). The metaphoric logic of musical motion and space. Ph.D. Dissertation, University of Oregon. Eitan, Z. (2007). Intensity contours and cross-dimensional interaction in music: Recent research and its implications for performance studies. Orbis Musicae, 14, 141–166. Eitan, Z., & Adler, M. (2008). Cross-domain mappings of pitch in Bach's Goldberg Variations. Unpublished raw data. Eitan, Z., & Granot, R. Y. (2006). How music moves: Musical parameters and images of motion. Music Perception, 23, 221–248. Eitan, Z., & Granot, R. Y. (2007). Intensity changes and perceived similarity: Inter-parametric analogies. Musicae Scientiae, Discussion Forum 4a, 99–133. Eitan, Z., Schupak, A., & Marks, L. E. (2008). Louder is higher: Cross-modal interaction of loudness change and vertical motion in speeded classification. In K. Miyazaki, Y. Hiraga, M. Adachi, Y. Nakajima, & M. Tsuzaki (Eds.), Proceedings of the 10th international conference on music perception and cognition (ICMP10) (pp. 2008). Adelaide: Causal Productions. Hubbard, T. L. (1996). Synesthesia-like mappings of lightness, pitch, and melodic interval. American Journal of Psychology, 109, 219–238. Lakoff, G., & Johnson, M. (1980a). Metaphors we live by. Chicago: University of Chicago Press. Lakoff, G., & Johnson, M. (1980b). The metaphorical structure of the human conceptual system. Cognitive Science, 4, 195–208. Lewkowicz, D. J., & Turkewitz, G. (1980). Cross-modal equivalence in early infancy: Auditory–visual intensity matching. Developmental Psychology, 6, 597–607. Lidji, P., Kolinsky, R., Lochy, A., & Morais, J. (2007). Spatial associations for musical stimuli: A piano in the head? Journal of Experimental Psychology: Human Perception and Performance, 33, 1189–1207. Marks, L. E. (1982). Bright sneezes and dark coughs, loud sunlight and soft moonlight. Journal of Experimental Psychology: Learning, Memory, and Cognition, 8, 177–193. Marks, L. E. (1987). On cross-modal similarity: Auditory-visual interactions in speeded discrimination. Journal of Experimental Psychology: Human Perception and Performance, 13, 384–394. Marks, L. E. (1996). On perceptual metaphors. Metaphor and Symbolic Activity, 11, 39–66. Marks, L. E. (2000). Synesthesia. In E. A. Cardeña, S. J. Lynn, & S. C. Krippner (Eds.), Varieties of anomalous experience: Phenomenological and scientific foundations. Washington, DC: American Psychological Association. Marks, L. E. (2004). Cross-modal interactions in speeded classification. In G. Calvert, C. Spence, & B. E. Stein (Eds.), Handbook of multisensory processes (pp. 85–106). Cambridge, Massachusetts: MIT Press. Marks, L. E., Hammeal, R. J., & Bornstein, L. B. (1987). Perceiving similarity and comprehending metaphor. Monographs of the Society for Research in Child Development, 52. Meier, B. P., & Robinson, M. D. (2004). Why the sunny side is up: Associations between affect and vertical position. Psychological Science, 15, 243–247. Meier, B. P., & Robinson, M. D. (2005). The metaphorical representation of affect. Metaphor and Symbol, 20, 239–257. Melara, R. D., & Marks, L. E. (1990). Interaction among auditory dimensions: Timbre, pitch, and loudness. Perception and Psychophysics, 482, 169–178. Melara, R. D., & O'Brien, T. P. (1987). Interaction between synesthetically corresponding dimensions. Journal of

Experimental Psychology: General, 116, 323–336. Merriam, A. (1964). The anthropology of music. Bloomington: Northwestern University Press. Mondloch, C. J., & Maurer, D. (2004). Do small white balls squeak? Pitchobject correspondences in young children. Cognitive, Affective, & Behavioral Neuroscience, 4, 133–136. Morton, E. S. (1977). On the occurrence and significance of motivationstructural rules in some bird and mammal sounds. American Naturalist, 111, 855–869. Morton, E. S. (1982). Grading, discreteness, redundancy, and motivationstructural rules. In D. E. Kroodsma, E. H. Miller, & H. Ouellet (Eds.), Acoustic communication in birds (pp. 182–212). New York: Academic Press. Nagy, W. (1974). Figurative patterns and redundancy in the lexicon. Ph.D. Dissertation, University of California at San Diego. Neuhoff, J. G., & McBeath, M. K. (1996). The Doppler illusion: The influence of dynamic intensity change on perceived pitch. Journal of Experimental Psychology: Human Perception and Performance, 22, 970–985. Neuhoff, J. G., McBeath, M. K., & Wanzie, W. C. (1999). Dynamic frequency change influences loudness perception: A central, analytic process. Journal of Experimental Psychology: Human Perception and Performance, 25, 1050–1059. Nichols, J. (1971). Diminutive consonant symbolism in western North America. Language, 47, 826–848. Nuckolls, J. B. (1999). The case for sound symbolism. Annual Review of Anthropology, 28, 225–252. Nygaard, L. N., Herold, D. S., & Namya, L. L. (2009). The semantics of prosody: Acoustic and perceptual evidence of prosodic correlates to word meaning. Cognitive Science, 33, 127–146. Ohala, J. (1982). The voice of dominance. Journal of the Acoustic Society of America, 72, S66. Ohala, J. (1994). The frequency code underlies the sound-symbolic use of voice pitch. In L. Hinton, J. Nichols, & J. J. Ohala (Eds.), Sound symbolism (pp. 325–347). Cambridge, UK: Cambridge University Press. Osgood, C. E., Succi, G. J., & Tannenbaum, P. H. (1957). The measurement of meaning. Urbana, IL: University of Illinois Press. Pratt, C. C. (1930). The spatial character of high and low tones. Journal of Experimental Psychology, 13, 278–285. Roffler, S. K., & Butler, R. A. (1968). Localization of tonal stimuli in the vertical plane. Journal of the Acoustical Society of America, 43, 1260–1265. Rusconi, E., Kwan, B., Giordano, B. L., Umilta, C., & Butterworth, B. (2005). Spatial representation of pitch height: The SMARC effect. Cognition, 20, 1–17. Scherer, K. R. (2003). Vocal communication of emotion: A review of research paradigms. Speech Communication, 40, 227–256. Scherer, K. R., & Bänziger, T. (2004). Emotional expression in prosody: A review and an agenda for future research. In Speech prosody 2004: Proceedings (pp. 359–366). Retrieved 28.05.09. Seitz, J. A. (2005). The neural, evolutionary, developmental, and bodily basis of metaphor. New Ideas in Psychology, 23, 74–95. Shintel, H., Nusbaum, H. C., & Okrent, A. (2006). Analog acoustic expression in speech communication. Journal of Memory and Language, 5, 167–177. Simon, J. R., Mewaldt, S. P., Acosta, E., & Hu, J. (1976). Processing auditory information: Interaction of two population stereotypes. Journal of Applied Psychology, 61, 354–358. Smith, L. B., & Sera, M. D. (1992). A developmental analysis of the polar structure of dimensions. Cognitive Psychology, 24, 99–142. Stevens, S. S. (1975). Psychophysics. New York: John Wiley. Stewart, L., Walsh, V., & Frith, U. (2004). Reading music modifies spatial mapping in pianists. Perception & Psychophysics, 662, 183–195. Stone, R. M. (1981). Toward a Kpelle conceptualization of music performance. Journal of African Folklore, 94, 188–206. Tamir, H., & Eitan, Z. (2007). Higher is faster, but for whom? Pitch register, tempo preferences, musical training and gender. Paper presented to the 8th Conference of the Society for Music Perception and Cognition (SMPC), Montreal, Canada, June 2007. Timble, O. C. (1934). Localization of sound in the anterior–posterior and vertical dimensions of ''auditory'' space. British Journal of Psychology, 24, 320–334. Tsur, R. (1992). What makes sound patterns expressive: The poetic mode of speech-perception. Durham, NC: Duke University Press. Wagner, Y. S., Winner, E., Cicchetti, D., & Gardner, H. (1981). ''Metaphorical'' mapping in human infants. Child Development, 52, 728–731. Walker, P., Brenner, G., Spring, J., Masttock, K., Slater, A., & Johnson, S. (in press). Preverbal infants' sensitivity to synaesthetic cross modality correspondences. Psychological Science.. Walker, P., & Smith, S. (1984). Stroop interference based on the synaesthetic qualities of auditory pitch. Perception, 13, 75–81. Walker, P., & Smith, S. (1985). Stroop interference based on the multimodal correlates of haptic size and auditory pitch. Perception, 14, 729–736. Walker, P., & Smith, S. (1986). The basis of Stroop interference involving the multimodal correlates of auditory pitch. Perception, 15, 491–496. Walsh, V. (2003). A theory of magnitude: Common cortical metrics of time, space and quantity. Trends of Cognitive Science, 7, 483–488. Zbikowski, L. (1998). Metaphor and music theory: Reflections from cognitive science. Music Theory Online, 4.

1 Bulletin of the Council for Research in Music Education Spring, 2003, No. 156 Effects of Auditory and Motor Mental Practice in Memorized Piano Performance Zebulon Highben Ohio State University Caroline Palmer McGill University Abstract The study examined the effects of two types of mental practice in learning to perform an unfamiliar piece: auditory and motor practice. Sixteen pianists performed unfamiliar music from notation during ten practice trials, with the presence or absence of auditory feedback and motor feedback (finger movements). Pianists were instructed to mentally practice the missing feedback. After the practice trials the pianists performed from memory under normal performance conditions. Errors during performance from memory showed significant effects of both motor and auditory feedback during practice. Comparisons with aural skills posttests indicated that pianists with strong aural skills were least affected in memory tests by removal of auditory feedback during practice. Pianists with high self-ratings of playing by ear scored higher on the aural skills test and performed better from memory in the absence of auditory feedback. These findings suggest that an accurate auditory image is important for successful performance from memory. Introduction Many studies have examined the factors that influence performers as they learn unfamiliar music. One of the most crucial factors is practice. Two major components of practice are the auditory information that performers receive in feedback about their productions, and the motor information they receive from kinesthetic movements associated with the execution of the performance (Finney & Palmer, 2003; Palmer & Meyer, 2000; Ross, 1985). Finney and Palmer (2003) showed that the absence of auditory feedback during practice of an unfamiliar piece did not affect pianists' accuracy while the music notation was in front of them, but the absence of auditory feedback during practice did impair their later performance from memory. Comparison of performance errors made by novice and skilled pianists suggested that novice pianists are more dependent on motor information (hand and finger positions) during learning than are skilled pianists (Palmer & Meyer, 2000). Ross' (1985) study of trombonists indicated that information received during normal practice conditions (rehearsal with both auditory and kinesthetic feedback present) allowed musicians to make corrections and adjustments to their performance. McPherson (1995, 1997) demonstrated positive correlations between musicians' ability to perform rehearsed music and their sight-reading, playing by ear, and improvising abilities. In addition to physical practice, mental practice can also influence performers as they learn unfamiliar music. Coffman (1990) described mental practice as "the covert or imaginary rehearsal of a skill without muscular movement or sound" and showed that pianists' mental practice improved their performance compared with no practice. Although several meta-analyses have compared effects of different types of mental practice (Driskell, Copper & Moran, 1994; Feltz & Landers, 1983), few studies have compared types of mental practice in the context of music performance. Pianists' mental practice with an auditory model showed advantages over mental practice alone (Lim & Lippman, 1991; Theiler & Lippman, 1995). Rubin-Rabson (1937)

showed that analytical pre-study of the score often aided performers' memorization of unfamiliar music; this analytical study may lead to auditory imagery or motor imagery (Lim & Lippman,1991). Mental practice may help musicians learning to perform unfamiliar music by facilitating the creation of an auditory and/or motoric image. Comparison of different practice conditions showed that listening to a performance is an effective aid to learning unfamiliar music (Rosenthal, 1984; Rosenthal, Wilson, Evans, & Greenwalt, 1988), and musicians given supplemental aural training instruction were more accurate in error detection than a control group without such training (Sheldon, 1998). Not all studies show a facilitation of mental practice, however; Rosenthal et al (1988) found that silent analysis (mental practice, without explicit instructions) followed by sightreading was not more effective than just sightreading. The purpose of this study is to contrast mental and physical practice, in terms of auditory and motor feedback, to determine their role in how pianists learn to perform unfamiliar music. Effects of mental practice were tested by replacing auditory or motor feedback with instructions to imagine the missing feedback: how the piece sounds, or how the finger movements feel, during practice. After performers practiced from a musical score, the score was removed, and pianists performed from memory under normal feedback conditions. Because meta-analyses of mental practice indicated individual differences in imagery abilities that may influence the efficacy of mental practice, we use a within-subjects design to compare physical and mental practice effects on each performers' memory. We also collected two independent measures of mental imagery ability, one for motor imagery and one for auditory imagery, for comparison with the effects of mental practice. Finally, pianists gave self-report measures of memorization, playing by ear, and sight-reading abilities. Method Participants. Sixteen adult pianists were recruited from the Columbus, OH community; all pianists were currently performing, and half were music majors at a large midwestern university. Pianists were informed they would be sight-reading and memorizing short musical excerpts at the time of recruitment. Participants had at least 6 years of individual piano instruction (mean = 10 years, range = 6 to 15). Stimulus Materials. Four musical pieces were composed for the experiment, based on the compositional style of early Baroque era organ works. Each piece was two measures long, in a different key signature (two were major and two were minor) and each consisted of 20 individual tones in the left and right hand parts. Figure 1 displays the four musical pieces, with initial fingering positions. For the Auditory Only practice condition, an auditory recording of each stimulus was created. The recordings had a tempo of 90 beats per minute with a concert piano timbre and metronomic timing (each quarternote duration = 625 ms). Two posttests were used to provide independent measures of auditory and motor imagery abilities. The auditory posttest was adapted from Wing's (1968) tests of aural skills. Participants were shown a single-line melody (7-10 pitches) and heard a similar melody, which was the same as the notated melody or had a difference of one pitch. The stimuli in the Wing tests did not control for the size of interval changes, however, which made some melodies easier than others. The stimuli were adjusted by making the one-note difference a change of 1-2 semitones; the total number of changes that moved up or down in pitch were balanced. Eight of the 12 melodies presented had a one-note difference. The 12 melodies were sounded over headphones, and subjects compared the notation and the aural stimuli in a same-different task. Each stimulus was sounded twice, and each had a quarter-note duration of 625 ms (total duration of each stimulus was between 2500 – 5000 ms). Participants were told to indicate if any of the printed notes were different from the ones sounded. The motor imagery posttest, motivated by work on infant imitation of motor movements (Gleissner, Meltzoff & Bekkering, 2000), contained a sequence of 7 pictures of the right hand with one highlighted finger in each picture, to correspond to a sequence of finger movements. Pianists were instructed to imagine and memorize the movement sequence while holding their hands in loose fists. Then the sequence was removed and a sequence of 7 letters was presented that corresponded to pitches on the keyboard (such as G C F E D E C). Pianists were told to memorize and perform this pitch sequence with the right hand on an electronic piano with no auditory feedback, and to indicate whether the sequence they performed was the same or different from the imagined sequence. There were 12 stimuli, 8 of which had a single movement (digit) difference between the imagined sequence and the performed sequence. The differences were always one digit away from the correct finger in the imagined sequence. Equipment. Pianists' performances were recorded in MIDI on a PC-monitored Roland RD600 digital keyboard using the "Standard Concert Piano 1" timbre. The presentation of Figure 1. Musical pieces composed for the experiment. 4 Mental Practice in Piano Performance auditory stimuli and the recording of pitch and timing information in the performances was controlled by computer. The pianists heard their performances through AKG-K270 Studio headphones. Design. Each experimental session consisted of practice conditions, during which the independent variables of auditory and motor feedback were manipulated, followed by test conditions, during which pianists performed from memory under normal performance conditions (auditory and motor feedback present). In the Normal practice condition, pianists moved their fingers on the keys and heard themselves play over headphones during practice. In the Motor Only practice condition, the pianists moved their fingers on the keys but auditory feedback was removed. In the Auditory Only practice condition, motor feedback was removed (the pianists held their fingers in loose fists) but auditory feedback was present (pianists heard a computer-generated recording of the piece). In the Covert practice condition, both motor feedback and auditory feedback were removed during practice (pianists held their fingers in loose fists and heard silence). Each subject participated in each practice condition with a different stimulus piece (practice condition was a within-subject variable). The ordering of the 4 practice conditions and order of stimuli were counterbalanced across subjects with a Latin-square design. Procedure. Pianists were first given a consent form and a questionnaire that examined their musical training and background. In a practice trial, pianists practiced a 1-measure musical excerpt (similar to the actual stimuli) 5 times under normal conditions (with normal motor and auditory feedback). Then the notation was removed and the pianists performed the piece 3 times from memory. The pianists then performed the four conditions of the experiment; in each condition, pianists received specific instructions and practiced each piece 10 times with the musical notation present. In the Normal practice condition, pianists were simply instructed to perform the piece. In conditions in which one or both types of feedback were removed, pianists were instructed to imagine the missing feedback. In the Motor Only practice condition, they were told that they would not be able to hear the performance as they moved their fingers on the keys, and to imagine what the piece would sound like; in this condition, the first chord of the piece was sounded at the beginning to serve as an auditory reference. In the Auditory Only practice condition, they were told to hold their hands and fingers in loose fists (to reduce any undesired motor movements) as they heard the piece and to imagine what the finger movements would feel like. In the Covert practice condition, both of these instructions were given; in addition, pianists were told to depress the foot pedal each time they began to imagine the piece's sound and movements (to supply a measure of amount of time spent imagining). After each practice stage ended, the notation was removed and pianists played the piece 4 times from memory. During these 4 test trials, the feedback conditions were always normal. Following the completion of the memory task, the two imagery posttests were administered. The motor imagery test was conducted first; pianists memorized the pictorial sequence of finger movements and used the electronic keyboard to "perform" (without auditory feedback) the subsequent letter sequences, and recorded their samedifferent answers on an answer sheet. Then the auditory imagery test was administered. On each trial, pianists listened to short sequences presented over headphones while reading music from notation, and indicated same-different on an answer sheet. The subjects were also asked to circle which pitch had changed in each example, if they were able. Finally, self-report measures were administered; pianists were asked to rate their sight-reading, memorizing, and playing by ear skills on a 10-point Likert scale (1=poor, 10 = excellent). Highben & Palmer 5 Results The number of pitch errors was detected in the MIDI-recorded performances with a computer error-detection algorithm (see Large, 1993; Palmer & van de Sande, 1993), and error scores were converted to percentage of total notes in the music that were correctly recalled. (Duration errors were rare, due perhaps to the rhythmic simplicity of the music.) Performance at the end of practice was examined first, to ensure that the pianists correctly learned the music; participants were included only if the last 4 trials in the Normal practice condition showed percent correct notes of more than 95% . An analysis of variance on the percent correctly recalled notes at test (from memory) by practice condition (Normal, Auditory Only, Motor Only, Covert) indicated a significant effect of practice condition, $F(3, 45) = 8.2$, $p < .01$. As shown in Figure 2, performance following Normal practice was best and Covert practice was worst. Tukey post-hoc tests indicated that performance in the Covert condition was significantly worse than in each of the other conditions ($p < .05$), and performance in the Normal condition was significantly better than in the Motor Only condition ($p < .05$). There was no significant difference between Motor Only and Auditory Only conditions, or between Normal and Auditory Only conditions. Presence of auditory or motor feedback at practice aided pianists' recall at test. Figure 2. Pianists' mean percentage of pitches correctly recalled from memory at test for each feedback practice condition. Pianists' scores on the auditory imagery posttest (perceiving melodic differences) and the motor imagery posttest (perceiving fingering differences), computed as percent correct responses, were examined next. A paired-sample t-test indicated that pianists' scores on the auditory imagery posttest (mean = 62.8 %) were significantly lower ($t(15) = -2.86$, $p < .05$) than their scores on the motor imagery posttest (mean = 81.8 %). There were no significant correlations between scores on the motor posttest and performance in any of the memory task conditions. The correlations between the auditory posttest scores and the memory task were significant; the auditory posttest scores correlated with pianists' recall scores in the memory task following the Covert 6 Mental Practice in Piano Performance condition ($r = .76$, $p < .01$) and the Motor Only condition ($r = .51$, $p < .05$), the two conditions in which auditory feedback was removed and pianists were instructed to imagine the sound. Pianists with higher auditory posttest scores were less affected by the removal of auditory feedback at practice than those with lower auditory posttest scores. To further examine the effects of aural skills on the memory task, the pianists were divided into 2 groups based on a median split of the auditory posttest scores: those with the 8 highest scores (mean = 80%) and 8 lowest scores (mean = 46%). Results in the memory task for the high and low scorers are shown in Figure 3. There was a significant interaction between aural skills level and the practice condition, $F(3, 42) = 2.72$, $p = .05$. As shown in Figure 3, pianists with lower aural skills scores performed worse from memory than pianists with higher scores, in the Motor Only and Covert practice conditions — the two conditions in which auditory feedback was removed. Aural skills aided pianists' ability to memorize music that was learned in the absence of sound. Finally, the pianists' self-ratings of sight-reading ability, ability to memorize, and playing by ear were compared with other performance measures. Only self-ratings of playing by ear (mean = 4.7 out of 10) correlated with other variables; pianists' ratings of playing by ear correlated significantly with their auditory posttest scores ($r = .71$, $p < .01$) and with performance from memory at test across all practice conditions ($r = .54$, $p < .01$). Self-ratings of memorizing skills (mean = 7.2) or sight-reading skills (mean = 6.8) did not correlate with any variables. Discussion Both auditory and motor forms of practice facilitated pianists' subsequent performance from memory of unfamiliar music. Removal of auditory or motor feedback at practice caused significant memory deficits in later performance, despite the presence of both types of feedback at test. Physical practice conditions (in which auditory and/ or motor feedback were present) led to better performance recall than conditions with Figure 3. Mean percentage of pitches correctly recalled from memory at test for each feedback practice condition, for pianists with higher (n=8) and lower (n=8) aural skills scores. Highben & Palmer 7 mental practice. Recall was best following normal practice conditions, and worst when both auditory and motor feedback were removed during practice. Furthermore, performers who had strong aural skills were least affected by the absence of auditory feedback during learning. Whereas previous studies demonstrated the overall efficacy of mental practice in music performance (Coffman, 1990; Ross, 1985), these findings suggest specifically that auditory forms of mental practice aid performers' learning of unfamiliar music. Good performance in the absence of feedback could be due to little or no reliance on that type of feedback, or alternatively, to good imagery skills that allow performers to use mental practice to "fill in" for the missing feedback. Several aspects of this study are consistent with the latter explanation. First, there was a strong correlation between individual performers' aural skills measures and their performance from memory following the absence of auditory feedback at practice (Motor Only and Covert conditions). This finding suggests that musicians with strong aural skills can successfully create an auditory image of that information during practice. Second, the correlations between pianists' subjective measures of their ability to play by ear, the objective aural skills measures, and performance from memory in all memory conditions support the notion that mental practice can help substitute for missing auditory information. The advantage of an accurate auditory image during mental practice is consistent with previous studies that document the increased efficacy of mental practice with an auditory model (Lim & Lippman, 1990; Rosenthal et al, 1988; Theiler & Lippman, 1995). Although the measures of performance from memory, aural skills, and self-ratings of playing by ear consistently point to the role of auditory mental practice, some aspects of the findings may be attributed to the design of the study. For example, all pianists were required to be capable of performing the music correctly by the end of practice, to ensure that the results reflected memory problems instead of learning problems. This criterion could account for the lack of effects of the motor imagery posttest, on which all of the pianists scored high. The test may have been too easy for the calibre of pianists who participated in the study. Novice performers may have weaker motor imagery abilities than the pianists included in this study (Palmer & Meyer, 2000). In addition, the method of practicing a piece ten times without stopping or correcting mistakes, although allowing consistent amounts of practice across individuals, was somewhat artificial. Despite these limitations, the study provides some insight into the relationship between practice and performance from memory. The significance of auditory feedback during practice for improving performance from memory suggests that performers should concentrate on the sound of the piece during practice rather than on the movements that create that sound. In other words, the goals of practice should match the goals of performance. Music performance is often depicted as both a cognitive and motoric task (Palmer & Meyer, 2000; Theiler & Lippman, 1995); from this perspective, it is not surprising to find effects of both auditory and motor mental practice in performance from memory. What is more surprising is the association documented here between individual performers' memory and imagery skills. The individual differences suggest that performers differ in their motor imagery and learning versus their auditory imagery and learning. Although most comparisons of mental practice methods rely on cross-group comparisons (see Driskell, Copper & Moran, 1994; Feltz & Landers, 1983), the presence of individual differences suggests that within-subject designs — those that allow comparisons across all conditions within individual performers — may be necessary for evaluating the relative efficacy of mental practice methods. Future studies may trace the development of auditory imagery and motor imagery abilities in performance. 8 Mental Practice in Piano Performance References

Coffman, D.D. (1990). Effects of mental practice, physical practice, and knowledge of results in piano performance. Journal of Research in Music Education, 38, 187-196. Driskell, J.E., Copper, C., & Moran, A. (1994). Does mental practice enhance music performance? Journal of Applied Psychology, 79, 481-492. Feltz, D.L, & Landers, D.M. (1983). The effects of mental practice on motor skill learning and performance: A meta-analysis. Journal of Sport Psychology, 5, 25-27. Finney, S.A. & Palmer, C. (2003). Auditory feedback and memory for music performance: Sound evidence for an encoding effect. Memory & Cognition, 31, 51-64. Gleissner, B., Meltzoff, A.N., & Bekkering, H. (2000). Children's coding of human action: cognitive factors influencing imitation in 3-year-olds. Developmental Science, 3(4), 405-414. Large, E.W. (1993).Dynamic programming for the analysis of serial behaviors. Behavior Research Methods, Instruments, and Computers, 25, 238-241 . Lim, S., & Lippman, L.G. (1991). Mental practice and memorization of piano music. Journal of General Psychology, 118, 21-30. McPherson, G.E. (1995). Five aspects of musical performance and their correlates. Bulletin of the Council for Research in Music Education, 127, 115-121. McPherson, G.E. (1997). Cognitive strategies and skill acquisition in musical performance. Bulletin of the Council for Research in Music Education, 133, 64-71. Norman, D.A. (1981). Categorization of action slips. Psychological Review, 88, 1-15. Palmer, C. (1997). Music performance. Annual Review of Psychology, 48, 115-138. Palmer, C., & Meyer, R.K. (2000). Conceptual and motor learning in music performance. Psychological Science, 11(1), 63-68. Palmer, C., & van de Sande, C. (1993). Units of knowledge in music performance. Journal of Experimental Psychology: Learning, Memory, & Cognition, 19, 457-470. Rosenthal, R.K. (1984). The relative effects of guided model, model only, guided only, and practice only on the accuracy of advanced instrumentalists' musical performance. Journal of Research in Music Education, 32, 265-273. Rosenthal, R.K, Wilson, M., Evans, M., & Greenwalt, L. (1988). Effects of different practice conditions on advanced instrumentalists' performance accuracy. Journal of Research in Music Education, 36, 250-257. Ross, S.L. (1985). The effectiveness of mental practice in improving the performance of college trombonists. Journal of Research in Music Education, 33, 221-230. Rubin-Rabson, G. (1937). The influence of analytical pre-study in memorizing piano music. Archives of Psychology, 20, 3-53. Sheldon, D.A. (1998). Effects of contextual sight-singing, and aural skills training on error-detection abilities. Journal of Research in Music Education, 46, 384-395. Theiler, A.M., & Lippman, L.G. (1995). Effects of mental practice and modeling on guitar and vocal performance. Journal of General Psychology, 122, 329-343. Wing, H.D. (1968). Tests of musical ability and appreciation (2nd Ed). Cambridge: Cambridge University Press. Acknowledgements Zebulon Highben, Department of Psychology, Ohio State University, and Caroline Palmer, Department of Psychology, McGill University. This research was supported in part by an Ohio State University Office of Research Scholarship to the first author, and by the Canada Research Chairs Program and NIMH Grant R01-45764 to the second author. We thank David Butler, Patricia Flowers, Timothy Gerber, and David Huron for comments, and Grant Baldwin and Steven Finney for assistance. Correspondence concerning this article can be addressed to Caroline Palmer at the Department of Psychology, McGill University, Montreal, H3A 1B1 Canada, or caroline.palmer@mcgill.ca.

1 Bulletin of the Council for Research in Music Education Spring, 2003, No. 156 Effects of Auditory and Motor Mental Practice in Memorized Piano Performance Zebulon Highben Ohio State University Caroline Palmer McGill University Abstract The study examined the effects of two types of mental practice in learning to perform an unfamiliar piece: auditory and motor practice. Sixteen pianists performed unfamiliar music from notation during ten

practice trials, with the presence or absence of auditory feedback and motor feedback (finger movements). Pianists were instructed to mentally practice the missing feedback. After the practice trials the pianists performed from memory under normal performance conditions. Errors during performance from memory showed significant effects of both motor and auditory feedback during practice. Comparisons with aural skills posttests indicated that pianists with strong aural skills were least affected in memory tests by removal of auditory feedback during practice. Pianists with high self-ratings of playing by ear scored higher on the aural skills test and performed better from memory in the absence of auditory feedback. These findings suggest that an accurate auditory image is important for successful performance from memory. Introduction Many studies have examined the factors that influence performers as they learn unfamiliar music. One of the most crucial factors is practice. Two major components of practice are the auditory information that performers receive in feedback about their productions, and the motor information they receive from kinesthetic movements associated with the execution of the performance (Finney & Palmer, 2003; Palmer & Meyer, 2000; Ross, 1985). Finney and Palmer (2003) showed that the absence of auditory feedback during practice of an unfamiliar piece did not affect pianists' accuracy while the music notation was in front of them, but the absence of auditory feedback during practice did impair their later performance from memory. Comparison of performance errors made by novice and skilled pianists suggested that novice pianists are more dependent on motor information (hand and finger positions) during learning than are skilled pianists (Palmer & Meyer, 2000). Ross' (1985) study of trombonists indicated that information received during normal practice conditions (rehearsal with both auditory and kinesthetic feedback present) allowed musicians to make corrections and adjustments to their performance. McPherson (1995, 1997) demonstrated positive correlations between musicians' ability to perform rehearsed music and their sight-reading, playing by ear, and improvising abilities. In addition to physical practice, mental practice can also influence performers as they learn unfamiliar music. Coffman (1990) described mental practice as "the covert or imaginary rehearsal of a skill without muscular movement or sound" and showed that pianists' mental practice improved their performance compared with no practice. Although several meta-analyses have compared effects of different types of mental practice (Driskell, Copper & Moran, 1994; Feltz & Landers, 1983), few studies have compared types of mental practice in the context of music performance. Pianists' mental practice with an auditory model showed advantages over mental practice alone 2 Mental Practice in Piano Performance (Lim & Lippman, 1991; Theiler & Lippman, 1995). Rubin-Rabson (1937) showed that analytical pre-study of the score often aided performers' memorization of unfamiliar music; this analytical study may lead to auditory imagery or motor imagery (Lim & Lippman,1991). Mental practice may help musicians learning to perform unfamiliar music by facilitating the creation of an auditory and/or motoric image. Comparison of different practice conditions showed that listening to a performance is an effective aid to learning unfamiliar music (Rosenthal, 1984; Rosenthal, Wilson, Evans, & Greenwalt, 1988), and musicians given supplemental aural training instruction were more accurate in error detection than a control group without such training (Sheldon, 1998). Not all studies show a facilitation of mental practice, however; Rosenthal et al (1988) found that silent analysis (mental practice, without explicit instructions) followed by sightreading was not more effective than just sightreading. The purpose of this study is to contrast mental and physical practice, in terms of auditory and motor feedback, to determine their role in how pianists learn to perform unfamiliar music. Effects of mental practice were tested by replacing auditory or motor feedback with instructions to imagine the missing feedback: how the piece sounds, or how the finger movements feel, during practice. After performers practiced from a musical score, the score was removed, and pianists performed from memory under normal feedback conditions. Because meta-analyses of mental practice indicated individual differences in imagery abilities that may influence the efficacy of mental practice, we use a within-subjects design to compare physical and mental practice effects on each performers' memory. We also collected two independent measures of mental imagery ability, one for motor imagery and one for auditory imagery, for comparison with the effects of mental practice. Finally, pianists gave self-report measures of memorization, playing by ear, and sight-reading abilities. Method Participants. Sixteen adult pianists were recruited from the Columbus, OH community; all pianists were currently performing, and half were music majors at a large midwestern university. Pianists were informed they would be sight-reading and memorizing short musical excerpts at the time of recruitment. Participants had at least 6 years of individual piano instruction (mean = 10 years, range = 6 to 15). Stimulus Materials. Four musical pieces were composed for the experiment, based on the compositional style of early Baroque era organ works. Each piece was two measures long, in a different key signature (two were major and two were minor) and each consisted of 20 individual tones in the left and right hand parts. Figure 1 displays the four musical pieces, with initial fingering positions. For the Auditory Only practice condition, an auditory recording of each stimulus was created. The recordings had a tempo of 90 beats per minute with a concert piano timbre and metronomic timing (each quarternote duration = 625 ms). Two posttests were used to provide independent measures of auditory and motor imagery abilities. The auditory posttest was adapted from Wing's (1968) tests of aural skills. Participants were shown a single-line melody (7-10 pitches) and heard a similar melody, which was the same as the notated melody or had a difference of one pitch. The stimuli in the Wing tests did not control for the size of interval changes, however, which made some melodies easier than others. The stimuli were adjusted by making the one-note difference a change of 1-2 semitones; the total number of changes that moved up or down in pitch were balanced. Eight of the 12 melodies presented had a one-note difference. The 12 melodies were sounded over headphones, and subjects compared the notation and the aural stimuli in a same-different task. Each stimulus was sounded twice, and each had a quarter-note duration of 625 ms (total duration of each stimulus was between 2500 – 5000 ms). Participants were told to indicate if any of the printed notes were different from the ones sounded. The motor imagery posttest, motivated by work on infant imitation of motor movements (Gleissner, Meltzoff & Bekkering, 2000), contained a sequence of 7 pictures of the right hand with one highlighted finger in each picture, to correspond to a sequence of finger movements. Pianists were instructed to imagine and memorize the movement sequence while holding their hands in loose fists. Then the sequence was removed and a sequence of 7 letters was presented that corresponded to pitches on the keyboard (such as G C F E D E C). Pianists were told to memorize and perform this pitch sequence with the right hand on an electronic piano with no auditory feedback, and to indicate whether the sequence they performed was the same or different from the imagined sequence. There were 12 stimuli, 8 of which had a single movement (digit) difference between the imagined sequence and the performed sequence. The differences were always one digit away from the correct finger in the imagined sequence. Equipment. Pianists' performances were recorded in MIDI on a PC-monitored Roland RD600 digital keyboard using the "Standard Concert Piano 1" timbre. The presentation of Figure 1. Musical pieces composed for the experiment. auditory stimuli and the recording of pitch and timing information in the performances was controlled by computer. The pianists heard their performances through AKG-K270 Studio headphones. Design. Each experimental session consisted of practice conditions, during which the independent variables of auditory and motor feedback were manipulated, followed by test conditions, during which pianists performed from memory under normal performance conditions (auditory and motor feedback present). In the Normal practice condition, pianists moved their fingers on the keys and heard themselves play over headphones during practice. In the Motor Only practice condition, the pianists moved their fingers on the keys but auditory feedback was removed. In the Auditory Only practice condition, motor feedback was removed (the pianists held their fingers in loose fists) but auditory feedback was present (pianists heard a computer-generated recording of the piece). In the Covert practice condition, both motor feedback and auditory feedback were removed during practice (pianists held their fingers in loose fists and heard silence). Each subject participated in each practice condition with a different stimulus piece (practice condition was a within-subject variable). The ordering of the 4 practice conditions and order of stimuli were counterbalanced across subjects with a Latin-square design. Procedure. Pianists were first given a consent form and a questionnaire that examined their musical training and background. In a practice trial, pianists practiced a 1-measure musical excerpt (similar to the actual stimuli) 5 times under normal conditions (with normal motor and auditory feedback). Then the notation was removed and the pianists performed the piece 3 times from memory. The pianists then performed the four conditions of the experiment; in each condition, pianists received specific instructions and practiced each piece 10 times with the musical notation present. In the Normal practice condition, pianists were simply instructed to perform the piece. In conditions in which one or both types of feedback were removed, pianists were instructed to imagine the missing feedback. In the Motor Only practice condition, they were told that they would not be able to hear the performance as they moved their fingers on the keys, and to imagine what the piece would sound like; in this condition, the first chord of the piece was sounded at the beginning to serve as an auditory reference. In the Auditory Only practice condition, they were told to hold their hands and fingers in loose fists (to reduce any undesired motor movements) as they heard the piece and to imagine what the finger movements would feel like. In the Covert practice condition, both of these instructions were given; in addition, pianists were told to depress the foot pedal each time they began to imagine the piece's sound and movements (to supply a measure of amount of time spent imagining). After each practice stage ended, the notation was removed and pianists played the piece 4 times from memory. During these 4 test trials, the feedback conditions were always normal. Following the completion of the memory task, the two imagery posttests were administered. The motor imagery test was conducted first; pianists memorized the pictorial sequence of finger movements and used the electronic keyboard to "perform" (without auditory feedback) the subsequent letter sequences, and recorded their samedifferent answers on an answer sheet. Then the auditory imagery test was administered. On each trial, pianists listened to short sequences presented over headphones while reading music from notation, and indicated same-different on an answer sheet. The subjects were also asked to circle which pitch had changed in each example, if they were able. Finally, self-report measures were administered; pianists were asked to rate their sight-reading, memorizing, and playing by ear skills on a 10-point Likert scale (1=poor, 10 = excellent). Highben & Palmer 5 Results The number of pitch errors was detected in the MIDI-recorded performances with a computer error-detection algorithm (see Large, 1993; Palmer & van de Sande, 1993), and error scores were converted to percentage of total notes in the music that were correctly recalled. (Duration errors were rare, due perhaps to the rhythmic simplicity of the music.) Performance at the end of practice was examined first, to ensure that the pianists correctly learned the music; participants were included only if the last 4 trials in the Normal practice condition showed percent correct notes of more than 95% . An analysis of variance on the percent correctly recalled notes at test (from memory) by practice condition (Normal, Auditory Only, Motor Only, Covert) indicated a significant effect of practice condition, $F(3, 45) = 8.2$, $p < .01$. As shown in Figure 2, performance following Normal practice was best and Covert practice was worst. Tukey post-hoc tests indicated that performance in the Covert condition was significantly worse than in each of the other conditions ($p < .05$), and performance in the Normal condition was significantly better than in the Motor Only condition ($p < .05$). There was no significant difference between Motor Only and Auditory Only conditions, or between Normal and Auditory Only conditions. Presence of auditory or motor feedback at practice aided pianists' recall at test. Figure 2. Pianists' mean percentage of pitches correctly recalled from memory at test for each feedback practice condition. Pianists' scores on the auditory imagery posttest (perceiving melodic differences) and the motor imagery posttest (perceiving fingering differences), computed as percent correct responses, were examined next. A paired-sample t-test indicated that pianists' scores on the auditory imagery posttest (mean = 62.8 %) were significantly lower ($t (15) = -2.86, p < .05$) than their scores on the motor imagery posttest (mean = 81.8 %). There were no significant correlations between scores on the motor posttest and performance in any of the memory task conditions. The correlations between the auditory posttest scores and the memory task were significant; the auditory posttest scores correlated with pianists' recall scores in the memory task following the Covert 6 Mental Practice in Piano Performance condition ($r = .76, p < .01$) and the Motor Only condition ($r = .51, p < .05$), the two conditions in which auditory feedback was removed and pianists were instructed to imagine the sound. Pianists with higher auditory posttest scores were less affected by the removal of auditory feedback at practice than those with lower auditory posttest scores. To further examine the effects of aural skills on the memory task, the pianists were divided into 2 groups based on a median split of the auditory posttest scores: those with the 8 highest scores (mean = 80%) and 8 lowest scores (mean = 46%). Results in the memory task for the high and low scorers are shown in Figure 3. There was a significant interaction between aural skills level and the practice condition, $F (3, 42) = 2.72, p = .05$. As shown in Figure 3, pianists with lower aural skills scores performed worse from memory than pianists with higher scores, in the Motor Only and Covert practice conditions — the two conditions in which auditory feedback was removed. Aural skills aided pianists' ability to memorize music that was learned in the absence of sound. Finally, the pianists' self-ratings of sight-reading ability, ability to memorize, and playing by ear were compared with other performance measures. Only self-ratings of playing by ear (mean = 4.7 out of 10) correlated with other variables; pianists' ratings of playing by ear correlated significantly with their auditory posttest scores ($r = .71, p < .01$) and with performance from memory at test across all practice conditions ($r = .54, p < .01$). Self-ratings of memorizing skills (mean = 7.2) or sight-reading skills (mean = 6.8) did not correlate with any variables. Discussion Both auditory and motor forms of practice facilitated pianists' subsequent performance from memory of unfamiliar music. Removal of auditory or motor feedback at practice caused significant memory deficits in later performance, despite the presence of both types of feedback at test. Physical practice conditions (in which auditory and/ or motor feedback were present) led to better performance recall than conditions with Figure 3. Mean percentage of pitches correctly recalled from memory at test for each feedback practice condition, for pianists with higher (n=8) and lower (n=8) aural skills scores. Highben & Palmer 7 mental practice. Recall was best following normal practice conditions, and worst when both auditory and motor feedback were removed during practice. Furthermore, performers who had strong aural skills were least affected by the absence of auditory feedback during learning. Whereas previous studies demonstrated the overall efficacy of mental practice in music performance (Coffman, 1990; Ross, 1985), these findings suggest specifically that auditory forms of mental practice aid performers' learning of unfamiliar music. Good performance in the absence of feedback could be due to little or no reliance on that type of feedback, or alternatively, to good imagery skills that allow performers to use mental practice to "fill in" for the missing feedback. Several aspects of this study are consistent with the latter explanation. First, there was a strong correlation between individual performers' aural skills measures and their performance from memory following the absence of auditory feedback at practice (Motor Only and Covert conditions). This finding suggests that musicians with strong aural skills can successfully create an auditory image of that information during practice. Second, the correlations between pianists' subjective measures of their ability to play by ear, the objective aural skills measures, and performance from memory in all memory conditions support the notion that mental practice can help substitute for missing auditory information. The advantage of an accurate auditory image during mental practice is consistent with previous studies that document the increased efficacy of mental practice with an auditory model (Lim & Lippman, 1990; Rosenthal et al, 1988; Theiler & Lippman, 1995). Although the measures of performance from memory, aural skills, and self-ratings of playing by ear consistently point to the role of auditory mental practice, some aspects of the findings may be attributed to the design of the study. For example, all pianists were required to be capable of performing the music correctly by the end of practice, to ensure that the results reflected memory problems instead of learning problems. This criterion could account for the lack of effects of the motor imagery posttest, on which all of the pianists scored high. The test may have been too easy for the calibre of pianists who participated in the study. Novice performers may have weaker motor imagery abilities than the pianists included in this study (Palmer & Meyer, 2000). In addition, the method of practicing a piece ten times without stopping or correcting mistakes, although allowing consistent amounts of practice across individuals, was somewhat artificial. Despite these limitations, the study provides some insight into the relationship between practice and performance from memory. The significance of auditory feedback during practice for improving performance from memory suggests that performers should concentrate on the sound of the piece during practice rather than on the movements that create that sound. In other words, the goals of practice should match the goals of performance. Music performance is often depicted as both a cognitive and motoric task (Palmer & Meyer, 2000; Theiler & Lippman, 1995); from this perspective, it is not surprising to find effects of both auditory and motor mental practice in performance from memory. What is more surprising is the association documented here between individual performers' memory and imagery skills. The individual differences suggest that performers differ in their motor imagery and learning versus their auditory imagery and learning. Although most comparisons of mental practice methods rely on cross-group comparisons (see Driskell, Copper & Moran, 1994; Feltz & Landers, 1983), the presence of individual differences suggests that within-subject designs — those that allow comparisons across all conditions within individual performers — may be necessary for evaluating the relative efficacy of mental practice methods. Future studies may trace the development of auditory imagery and motor imagery abilities in performance.

References

Coffman, D.D. (1990). Effects of mental practice, physical practice, and knowledge of results in piano performance. Journal of Research in Music Education, 38, 187-196. Driskell, J.E., Copper, C., & Moran, A. (1994). Does mental practice enhance music performance? Journal of Applied Psychology, 79, 481-492. Feltz, D.L, & Landers, D.M. (1983). The effects of mental practice on motor skill learning and performance: A meta-analysis. Journal of Sport Psychology, 5, 25-27. Finney, S.A. & Palmer, C. (2003). Auditory feedback and memory for music performance: Sound evidence for an encoding effect. Memory & Cognition, 31, 51-64. Gleissner, B., Meltzoff, A.N., & Bekkering, H. (2000). Children's coding of human action: cognitive factors influencing imitation in 3-year-olds. Developmental Science, 3(4), 405-414. Large, E.W. (1993).Dynamic programming for the analysis of serial behaviors. Behavior Research Methods, Instruments, and Computers, 25, 238-241 . Lim, S., & Lippman, L.G. (1991). Mental practice and memorization of piano music. Journal of General Psychology, 118, 21-30. McPherson, G.E. (1995). Five aspects of musical performance and their correlates. Bulletin of the Council for Research in Music Education, 127, 115-121.

McPherson, G.E. (1997). Cognitive strategies and skill acquisition in musical performance. Bulletin of the Council for Research in Music Education, 133, 64-71. Norman, D.A. (1981). Categorization of action slips. Psychological Review, 88, 1-15. Palmer, C. (1997). Music performance. Annual Review of Psychology, 48, 115-138. Palmer, C., & Meyer, R.K. (2000). Conceptual and motor learning in music performance. Psychological Science, 11(1), 63-68. Palmer, C., & van de Sande, C. (1993). Units of knowledge in music performance. Journal of Experimental Psychology: Learning, Memory, & Cognition, 19, 457-470. Rosenthal, R.K. (1984). The relative effects of guided model, model only, guided only, and practice only on the accuracy of advanced instrumentalists' musical performance. Journal of Research in Music Education, 32, 265-273. Rosenthal, R.K, Wilson, M., Evans, M., & Greenwalt, L. (1988). Effects of different practice conditions on advanced instrumentalists' performance accuracy. Journal of Research in Music Education, 36, 250-257. Ross, S.L. (1985). The effectiveness of mental practice in improving the performance of college trombonists. Journal of Research in Music Education, 33, 221-230. Rubin-Rabson, G. (1937). The influence of analytical pre-study in memorizing piano music. Archives of Psychology, 20, 3-53. Sheldon, D.A. (1998). Effects of contextual sight-singing, and aural skills training on error-detection abilities. Journal of Research in Music Education, 46, 384-395. Theiler, A.M., & Lippman, L.G. (1995). Effects of mental practice and modeling on guitar and vocal performance. Journal of General Psychology, 122, 329-343. Wing, H.D. (1968). Tests of musical ability and appreciation (2nd Ed). Cambridge: Cambridge University Press. Acknowledgements Zebulon Highben, Department of Psychology, Ohio State University, and Caroline Palmer, Department of Psychology, McGill University. This research was supported in part by an Ohio State University Office of Research Scholarship to the first author, and by the Canada Research Chairs Program and NIMH Grant R01-45764 to the second author. We thank David Butler, Patricia Flowers, Timothy Gerber, and David Huron for comments, and Grant Baldwin and Steven Finney for assistance. Correspondence concerning this article can be addressed to Caroline Palmer at the Department of Psychology, McGill University, Montreal, H3A 1B1 Canada, or caroline.palmer@mcgill.ca.

https://doi.org/10.1177/10298649211062724 Musicae Scientiae 1–14 DOI: 10.1177/10298649211062724 journals.sagepub.com/home/msx Auditory and visual mental imagery in musicians and non-musicians Francesca Talamini Julia Vigl University of Innsbruck, Austria Elizabeth Doerr Massimo Grassi Barbara Carretti University of Padua, Italy Abstract Mental imagery plays an important role in various contexts of life, involving cognitive resources such as memory, learning, spatial representation, and reasoning. The vividness of mental images depends on different factors, including personal expertise in a certain field. For instance, musicians have been found to possess better auditory imagery abilities than non-musicians for both musical and nonmusical sounds. Only a few studies have tried to find out if this advantage is selective for auditory stimuli, however, with contradictory results so far (i.e., some studies supporting an advantage for mental imagery in general and some supporting an advantage for auditory mental imagery in particular). This study therefore investigated auditory and visual mental imagery in individuals with and without formal musical training. Thirty-six formally trained musicians, 33 self-taught musicians, and 33 non-musicians completed two questionnaires assessing the vividness of their auditory and visual mental imagery. They also completed measures of aptitude for music and general cognitive abilities. Both groups of musicians reported greater vividness of auditory (non-musical) imagery, but not visual imagery, than non-musicians. Thus musical experience, regardless of the type of training undergone by musicians, is linked to superior self-reported auditory mental imagery for everyday sounds, but not mental imagery in

general. Keywords professional musicians, self-taught musicians, everyday sounds, individual differences, sensory modality A mental image is an internally generated representation of an object, event, or sensation. In contrast to perception, it occurs when perceptual information is accessed from memory and is often described as “seeing with the mind’s eye” or “hearing with the mind’s ear” (Kosslyn et al., Corresponding author: Francesca Talamini, Department of Psychology, University of Innsbruck, Innrain 52, 6020 Innsbruck, Austria. Email: francesca.talamini@uibk.ac.at 1062724 MSX0010.1177/10298649211062724Musicae ScientiaeTalamini et al. research-article2021 Article 2 Musicae Scientiae 00(0) 2001, p. 635), depending on the sensory modality. Mental images can be internally generated as representations of things, experiences, or scenes (Schifferstein, 2009) allowing people to recreate the past and simulate the future (Moulton & Kosslyn, 2009). Here, the term vividness refers to how clearly and realistically a mental image is experienced (McAvinue & Robertson, 2007). Mental imagery plays a functional role in various contexts involving cognitive resources such as memory, learning, spatial representation, and reasoning. For instance, imagery is an effective method for promoting adherence to medical advice (Liu & Park, 2004); indeed, people who imagine implementing specific health procedures are more likely to perform them accurately and consistently in real life, even for extended periods of time. Mental imagery is also related to increased involvement in planned behavior and promotion of task engagement (Renner et al., 2019; Vasquez & Buehler, 2007), more effective learning outcomes (Guarnera et al., 2019), positive mood training (Holmes et al., 2006), improved sports performance (Mizuguchi et al., 2012), and motor control (Isaac & Marks, 1994); it also plays a major role in the domain of music (Aleman et al., 2000; Brochard et al., 2004; Brown & Palmer, 2013). Although many studies focus on visual mental imagery, which is also the modality in which the highest vividness ratings are obtained (Schifferstein, 2009), it is known from the literature that mental imagery can take all other kinds of sensory forms (e.g., tactile, olfactory, motor, and auditory, see Andrade et al., 2014). The vividness of mental images depends not only on the sensory and affective qualities of the imagined stimulus (Bywaters et al., 2004), or the capacity of short- and long-term memory systems (Baddeley & Andrade, 2000), but also on expertise and indeed personal experience in different areas. For example, Isaac and Marks (1994) showed that specific groups of experts, such as elite athletes, physical education students, air traffic controllers, and pilots, reported more vivid mental imagery than matched controls, especially in certain domains (e.g., physical education students provided higher vividness ratings for motor imagery than students in other academic disciplines; similarly, athletes and air traffic controllers reported more vivid motor and visual imagery than control groups). Mental imagery in musicians Musicians are often regarded as a model for studying brain and behavioral changes after prolonged and intense training (e.g., Münte et al., 2002; Schlaug, 2001). Music training makes high cognitive demands, as it requires information from different sources to be integrated, and appropriate motor actions to be planned and executed (Sergent, 1993). For a musician to plan any musical performance, they must be able to create a mental representation and to imagine a desired interpretation (Holmes, 2005). These mental representations are mostly auditory musical images generated in the anticipation of actual auditory feedback. In other words, learning to use imagery is part of a musician’s training, whether auditory, such as imagining a pitch, or not, such as imagining a movement. For example, singers generate auditory images of their performance by imagining the exact notes they will sing (Keller & Koch, 2008), and pianists are able to imagine their dynamics and articulation (e.g., staccato, legato) while they are performing (Bishop et al., 2013b). Musicians can thus use musical mental imagery as part of their typical learning and performing routines, and when reading notated music silently (Bailes, 2006; Bishop et al., 2014; Brodsky et al., 2008; Gregg et al., 2008). Furthermore, musicians can use mental imagery to help them create anticipatory images enabling them to plan actions and execute movements (Keller, 2012). In this way, they can gain an enhanced expressive and interpretive understanding of the music they are to perform (Connolly & Talamini et al. 3 Williamon, 2004); they can also improve their creativity in relation to composing music (Wong & Lim, 2017). In ensembles, interpersonal coordination is facilitated by mental imagery, as musicians simulate their own actions and the actions of the other musicians during the performance (Keller & Appel, 2010; Pecenka & Keller, 2009). It is therefore not surprising that some studies have found that musicians possess better musical and auditory imagery abilities than non-musicians, either at the behavioral level or at the level of brain activity (Aleman et al., 2000; Bishop et al., 2013a; Herholz et al., 2008). This advantage has mostly been observed in objective assessments of mental imagery. Examples of these assessments include tasks in which imagery is needed to obtain the correct solution to a problem (e.g., Aleman et al., 2000; Brochard et al., 2004). For example, musicians were found to perform better than non-musicians when asked to imagine and compare the pitches associated with two different lyrics of a well-known song. They also performed better in another version of the same task using non-musical stimuli. Yet there was no difference between the performance of musicians and non-musicians on a task involving the visualization and mental comparison of objects (Aleman et al., 2000). By contrast, Brochard and colleagues (2004) observed that musicians performed better than non-musicians on an objective visual imagery task (i.e., maintaining a mental representation of a visual cue, after it has disappeared, to facilitate the localization of a target stimulus). Janata and Paroo (2006) found a positive relationship between years of music training and acuity of the auditory image of pitch and tempo, suggesting a possible causal role for music training. Another common approach to the investigation of mental imagery is the use of self-report questionnaires such as the Vividness of Visual Imagery Questionnaire (VVIQ, Marks, 1973), the Plymouth Sensory Imagery Questionnaire (Andrade et al., 2014), the Movement Imagery Questionnaire (Hall & Martin, 1997), and the Spontaneous Use of Imagery Scale (Nelis et al., 2014). Little research using these questionnaires compares musicians and non-musicians, however. Exceptions include the work of Campos and Fuentes (2016), who found that music students reported greater vividness and clarity of cutaneous, kinesthetic, gustatory, visual, and auditory imagery than students of other subjects. Di Nuovo and Angelica (2016) found that expert musicians reported more vivid motor mental images than their untrained counterparts. In another study investigating musicians only, however, no relationship between musical experience and self-reported vividness of visual imagery was observed (Clark & Williamon, 2012). These findings raise the question as to whether musicians possess enhanced general mental imagery abilities, or whether these abilities are only for auditory stimuli. The evidence supporting an advantage for musicians in relation to general mental imagery, from studies in which behavioral rather than self-report data were collected, is contradictory (e.g., Aleman et al., 2000; Brochard et al., 2004). Also, the advantage of musicians in auditory imagery tasks, particularly when musical stimuli are presented, might be explained by their use of mental imagery when practicing (Gregg et al., 2008), by the potentially positive effects of musical training on cognitive functions (for a review, see Swaminathan & Schellenberg, 2019), or by their more effective processing of imagery representations in auditory cortical areas (Aleman et al., 2000). Objectives of the study Because the results of previous studies were contradictory, we wanted to look more closely at potential differences between musicians and non-musicians in terms of their musical auditory and visual imagery, specifically in relation to its vividness, as reported by participants in the research. Musical imagery was excluded as it is often part of musicians' training and might also be associated with movement (i.e., playing) and visual imagery (e.g., for music 4 Musicae Scientiae 00(0) notation). Two self-report questionnaires assessing the vividness of auditory and visual imagery were administered to participants with different levels of musical skill: formally trained musicians, self-taught musicians, and non-musicians. Comparing formally trained and self-taught musicians could help determine whether musicians' superior self-reported imagery abilities are linked to formal music training or music making more generally. A link to formal music training is suggested by evidence that formally trained musicians have advantages in some aspects of auditory processing (Zendel & Alexander, 2020). Furthermore, trained musicians scored higher than self-taught musicians for creativity in musical and verbal domains (Palmiero et al., 2020), and links have been identified between creativity and mental imagery (Palmiero et al., 2011). To assess auditory mental imagery, we used a new self-report questionnaire modeled on the VVIQ (Marks, 1973), the Vividness of Auditory Imagery Questionnaire (VAIQ) (see Supplemental Appendix), as our purpose was to use comparable measures of auditory and visual mental imagery for everyday objects and sounds. The VVIQ was chosen as a model because it is a short test with good reliability, already used as a model for other imagery questionnaires such as the Vividness of Movement Imagery Questionnaire (VMIQ, Isaac et al., 1986). Although other scales also assess vividness of auditory imagery, we believed that they were not fully comparable to the VVIQ. The Bucknell Auditory Imagery Scale (BAIS; Halpern, 2015), for example, has a different structure and many of its items refer to music, so it was not suitable for our purposes. Musicians are known to have an advantage in music imagery (Brodsky et al., 2003) and, as we aimed to compare auditory and visual imagery in musicians and non-musicians, we did not want the questionnaire to include items directly linked to imagery for music. Other multisensory scales such as the Plymouth Sensory Imagery Questionnaire (Andrade et al., 2014) and the Questionnaire on Mental Imagery (Sheehan, 1967) include only five items for each sensory modality and, as we aimed to compare only two sensory modalities, we wanted the questionnaire to include more items for each one so as to improve its reliability, especially as our sample of participants was relatively small (Marsh et al., 1998). We also measured participants' aptitude for music1 using the shortest version of the Profile of Music Perception Skills (Mini-PROMS; Zentner & Strauss, 2017), to assess whether auditory imagery can be linked to aptitude regardless of music training and/or musical activity, and their cognitive abilities using two subtests of the Wechsler Adult Intelligence Scale (WAIS-IV; Wechsler, 2008, see "Materials" section for further details). Method Participants One hundred two young adults participated in the study. The mean age was 22.7years (min=19, max=33). There were 35 females and 67 males. Participants were (1) 36 formally trained musicians (conservatory students, music school students, and/or professional musicians); (2) 33 self-taught musicians who, as children, had had fewer than 2years' music lessons other than at school, and reported being unable to read music notation; and (3) 33 self-identified non-musicians who, as children, had had fewer than 2years' music lessons other than at school. The protocol was approved by the ethics committee of the University of Padua. Demographic details and scores for the WAIS-IV subtests are presented separately for each group in Table 1. Formally trained musicians had slightly more years of training and/or musical activity than self-taught musicians, and reported practicing for substantially more hours per week than the self-taught musicians. Talamini et al. 5 Materials Mental imagery questionnaires VVIQ. The original scale (Marks, 1973) was used because, unlike a more recent revision, its translation into Italian (Antonietti & Crespi, 1995) has been validated. Minor differences between the psychometric properties of the original and the revision are not of concern (Campos & Pérez-Fabello, 2009). It has 16 items in four groups, each group representing an object to be imagined: a familiar person, the rising sun, a familiar storefront, and a countryside. Each of the four items in a group refers to a specific characteristic of the object, for example, the color of the familiar person's clothes, the contour of their face and the body, their walking pace, and their posture. Participants are asked to rate the vividness of each image using a 5-point scale from 5 (perfectly clear and vivid as normal seeing) to 1 (no image at all, you only "know" that you are thinking of an object). In the original version, the scale ranges from 1 (perfectly clear and vivid as normal seeing) to 5 but the Italian translation uses a reversed scale. Participants must complete the questionnaire twice, once with their eyes open and once with their eyes closed, having reread each item. VAIQ. This scale was modeled on the VVIQ, requiring participants to form auditory, nonmusical images, and rate their vividness. It also has 16 items in four groups, each group representing an everyday auditory object to be imagined: the sound of a toothbrush when brushing one's teeth, a child laughing, church bells, a thunderstorm. Again, participants are asked to rate the vividness of each image using a 5-point scale from 5 (Perfectly realistic, as vivid as hearing it) to 1 (no image at all, I only "know" that I am thinking of an object); again, participants must complete the questionnaire twice, once with their eyes open and once with their eyes closed, having re-read each item. Administration of the VAIQ questionnaire to 149 psychology students showed that it has good reliability (Cronbach's α=.85). Factor analysis with varimax orthogonal rotation yielded one factor explaining 29.53% of the variance, in line with a previous study of the Spanish translation of the VVIQ (Campos et al., 2002). Corrected item total correlations were carried out for the pairs of items to which participants had responded with their eyes open and shut (see Supplemental Appendix, Table S1). Table 1. Age, education, performance (raw scores) on the WAIS-IV Visual Puzzles and Vocabulary subtests, music training and/or musical activity, aptitude for music, and hours of practice per week. Formally trained musicians n=36 (13 females) Self-taught musicians n=33 (4 females) Non-musicians n=33 (18 females) p value Age (years) 22.7 (2.1) 23.2 (2.1) 22.1 (1.6) .137 Education (years) 16.3 (1.9) 16.5 (1.6) 15.9 (1.4) .333 Visual Puzzles (max. score 26) 17.53 (4.34) 17.48 (3.72) 15.42 (3.1) .037 Vocabulary (max. score 57) 45.75 (8.14) 47.18 (5.97) 44.76 (6.03) .354 Music training and/or musical activity (years) 11.8 (3.7) 10.2 (2.3) <2.0 .032 Weekly practice (hr) 14.2 (9.2) 7.2 (4.9) None <.001 Mini-PROMS total score 52.61 (8.05) 49.06 (6.78) 38.60 (7.33) .037 ANOVA: analysis of variance; WAIS: Wechsler Adult Intelligence Scale; PROMS: Profile of Music Perception Skills. M (SD). In the right-most column, the p value represents the results of t tests and ANOVAs as appropriate; significant values are shown in bold. Mini-PROMS. This online test was used to test all participants' aptitude for music. It has four subtests investigating different aspects of music perception. In each subtest, the participant listens twice to a standard stimulus, followed by a comparison stimulus and is asked to judge whether the comparison stimulus is the same as or different from the standard stimulus. The Melody subtest assesses the participant's ability to recognize whether or not two short melodies are the same. The Tuning subtest assesses the ability to recognize if one of the notes of a chord in the comparison stimulus is mistuned. The Accent subtest assesses the ability to recognize whether or not two short series of accentuated rhythmic clicks are the same, and the Tempo subtest assesses the ability to recognize whether a synthetic rhythmic structure or a recorded sample of music is being played for the second time faster or slower. WAIS-IV. Two subtests of this scale were administered to test participants' general cognitive abilities; specifically, their nonverbal reasoning and semantic skills, as these could help to explain individual differences in mental abilities, including imagery (Shaw & DeMers, 1986). The Visual Puzzles subtest was used to test participants' nonverbal reasoning. This is a timed test in which the participant must look at a series of reference figures and, for each one, choose three out of five elements of the figure that can be combined to recreate it. A score of 1 is awarded for each correct response. Level of difficulty increases figure by figure, and the test ends when the participant has been awarded three consecutive scores of 0. The Vocabulary subtest was used to test participants' semantic skills. The researcher reads a list of words aloud, one at a time, and the participant has to provide a short definition of each one. Scores of 2 are awarded for definitions that are complete, 1 for those that are incomplete, and 0 for those that are incorrect. Again, the level of difficulty increases word by word, and the test ends when the participant has been awarded three consecutive scores of 0. Apparatus and procedure The Mini-PROMS is administered online2 and participants took the test inside in a single-walled standard audiology booth (IAC Acoustics) wearing a pair of Sennheiser HD 580 headphones. The sound level at which the test was administered was comfortable and fixed for all participants. First, participants signed a written consent form, provided demographic information about themselves, and completed a standard questionnaire about musical experience (e.g., training, type of instrument, listening to music). Second, the two subtests of the WAIS-IV were administered. Third, participants completed the mental imagery questionnaires, in counterbalanced order, such that half of the participants completed the VVIQ first followed by the VAIQ and half completed the VAIQ first followed by the VVIQ. Fourth, they took the Mini-PROMS test. Finally, they completed a questionnaire about their musical habits such as listening to music and dancing, and the two groups of musicians responded to additional questions about their music training and/or musical activity (e.g., years of training, hours of weekly practice, type of instrument played). Analyses We used one-way analyses of variance (ANOVAs) to test the effects of group (formally trained and self-taught musicians, and non-musicians) on scores for the WAIS-IV subtests and the MiniPROMS, representing cognitive abilities and aptitude for music respectively. We computed the sum of all the ratings for all the items in the VVIQ and VAIQ for each participant, as there were no missing data, and explored associations between cognitive abilities, aptitude for music, and vividness of auditory and visual imagery using Pearson's r. We conducted a repeated-measures Talamini et al. 7 ANOVA to test the effects of group, modality (auditory vs. visual imagery), and condition (eyes open vs. eyes closed) on auditory and visual imagery. We adjusted p values for false discovery rate (FDR; Benjamini & Hochberg, 1995). Results Cognitive abilities A one-way ANOVA revealed a statistically significant effect of group on scores for the Visual Puzzles subtest of the WAIS-IV, $F(2, 99)=3.39$, $p=.037$, $\eta2=.06$. Post hoc tests showed that formally trained musicians performed better than non-musicians, $t(63)=2.32$, $p=.035$, $d=0.56$, and self-taught musicians performed better than non-musicians, $t(62)=2.44$, $p=.035$, $d=0.60$, but there was no significant difference between the scores of formally trained musicians and selftaught musicians, $t(67)=0.04$, $p=.965$, $d=0.01$ (see Table 1). There was no significant effect of group on scores for the Vocabulary subtest, $F(2, 99)=1.05$, $p=.354$, $\eta2=.02$. Aptitude for music A one-way ANOVA revealed a statistically significant effect of group on total scores for the Mini-PROMS test, $F(2, 99)=32.17$, $p<.001$, $\eta2=.39$, such that formally trained musicians ($M=52.61$, $SD=8.05$) outperformed non-musicians ($M=38.70$, $SD=7.33$), $t(67)=7.51$, $p<.001$, $d=1.81$, self-taught musicians ($M=49.06$, $SD=6.78$) outperformed non-musicians, $t(64)=5.96$, $p<.001$, $d=1.47$, and formally trained musicians outperformed self-taught musicians, $t(66)=1.99$, $p=.05$, $d=0.48$. Associations between years of music training and/or musical activity, hours of weekly practice, cognitive abilities, aptitude for music, and vividness of auditory and visual imagery As shown in Table 2, significant but weak correlations were found between years of music training and/or musical activity and VAIQ score, hours of weekly practice and VAIQ score, and hours of weekly practice and VVIQ score with eyes open. Weak correlations were also found between scores for the Melody and Tuning subtests of the Mini-PROMS and the VAIQ and the VVIQ. There were no correlations between cognitive abilities and visual or auditory imagery so scores on the Visual Puzzles and Vocabulary subtests of the WAIS-IV were not included as covariates in subsequent analyses. Auditory and visual imagery A repeated-measures ANOVA revealed a significant effect of condition, $F(1, 99)=43.23$, $p<.001$, $\eta2=.29$, such that images were reported to be more vivid when participants' eyes were closed ($M=123.5$, $SD=18.59$) than when they were open ($M=114.5$, $SD=18.26$). There was no significant interaction between modality (visual vs.

auditory) and condition, $F(1, 99)=0.80$, $p=.372$, but there was a significant interaction between modality and group, $F(2, 99)=4.20$, $p=.018$, $\eta 2=.08$, illustrated in Figure 1. Post hoc tests showed that formally trained musicians scored significantly higher than non-musicians for auditory imagery, $t(67)=2.54$, $p=.0393$, $d=0.61$, but not visual imagery, $t(67)=1.28$, $p=.310$, $d=0.31$. Similarly, self-taught musicians scored significantly higher for auditory imagery, $t(64)=3.55$, $p=.004$, $d=0.87$, but 8 Musicae Scientiae 00(0) not visual imagery, $t(59)=1.58$, $p=.240$, $d=0.39$. There were no significant differences between the scores of formally trained and self-taught musicians for auditory imagery, $t(67)=-0.9$, $p=.445$, $d=0.22$, or visual imagery, $t(64)=-0.42$, $p=.679$, $d=0.10$. Discussion In this study, we assessed auditory and visual mental imagery in three groups of participants with different levels of musical skill: formally trained musicians, self-taught musicians, and Table 2. Pearson Correlations between Years of Music Training and/or Musical Activity, Hours of Weekly Practice, Mini-PROMS Scores, Visual Puzzles, Vocabulary, and VVIQ and VAIQ. VVIQ EO VVIQ EC VAIQ EO VAIQ EC Music training and/or musical activity (years) .03 .17 .20* .26** Practice hours .07 .22* .21* .30** Mini-PROMS Melody .18 .23* .30** .24* Mini-PROMS Tuning .08 .23* .20* .24* PROMS Tempo .10 .14 .17 .18 Mini-PROMS Accent .11 .13 .17 .18 Mini-PROMS Total .15 .22* .27** .27** Visual Puzzles −.03 −.06 .04 .04 Vocabulary .10 .10 .12 .07 VVIQ: Vividness of Visual Imagery Questionnaire; EO: eyes open; EC: eyes closed; VAIQ: Vividness of Auditory Imagery Questionnaire; PROMS: Profile of Music Perception Skills. Data representing music training and/or musical activity and hours of weekly practice were analyzed only for the two groups of musicians, as non-musicians had less than 2 years of music training. *p<.05, **p<.01. Figure 1. Self-report scores in the mental imagery questionnaires (sum of eyes open and eyes closed conditions): the higher the value, the higher the vividness reported. Note: For each modality, the score for the opened eyes condition and closed eyes condition was combined, to evidence the interaction between group and modality. The maximum score obtainable is 160. Inside each box, the horizontal line indicates the median. The red diamond represents the mean. The edges of the box represent the 25th and the 75th percentiles. The whiskers are the interquartile range (i.e., Q3–Q1) augmented by 50%. The black dots are outliers. Talamini et al. 9 non-musicians. We also assessed aptitude for music objectively using the Mini-PROMS. The results indicate that both formally trained and self-taught musicians reported more vivid mental images than non-musicians, but only for auditory objects. The differences between the scores of musicians and non-musicians representing vividness of auditory imagery were more striking for self-taught than formally trained musicians, as type of correction for multiple comparisons affected the p value of the result of the comparison between formally trained and non-musicians. However, this comparison was likely to have been affected by two outliers in the group of formally trained musicians (see Figure 1); when these observations were removed, the difference between groups remained consistently significant across different types of multi-comparison correction. There were no significant differences between the scores of the three groups representing vividness of visual imagery, a result in line with the previous finding that musicians have better mental imagery abilities than non-musicians, but only when they are asked to create auditory images (Aleman et al., 2000). Note that in the study by Aleman and colleagues (2000), however, a different paradigm was used to test mental imagery, with three conditions: musical (participants were asked to compare pitches), non-musical auditory (participants were asked to imagine the acoustic characteristics of everyday sounds), and visual (participants were asked to visualize the forms of objects). This would seem to strengthen the validity of the self-report questionnaires used in this study as measures of mental imagery ability. Nevertheless, there were no significant differences between the scores of the formally trained and self-taught musicians representing vividness of auditory imagery, suggesting that this ability is not necessarily associated with formal musical training only but, more generally, with musical activity. Moreover, years of musical activity (regardless of whether the individual has undergone formal training) could also play a role in the self-reported ability to imagine auditory stimuli. If so, this role is likely to be small, however, as we found only a weak correlation between years of music training and/or musical activity and scores on the VAIQ. This is in line with the finding in previous research by Pfordresher and Halpern (2013) that music training was correlated only weakly with a measure of auditory imagery. Note that the BAIS, a scale more oriented toward music, was used in this study. A comparison of formally trained and self-taught musicians with this scale could show a different pattern of results. (e.g., different imagery abilities between formally trained and self-taught musicians). Yet we cannot exclude the possibility that factors besides playing a musical instrument are responsible for the higher scores of our two groups of musicians on measures of auditory imagery. We found, for example, that both groups of musicians reported listening to music more often than non-musicians. In future research, the role of listening to music in auditory imagery should be clarified. This could be achieved by asking non-musicians about their listening behaviors. Finally, the vividness of auditory imagery seems also not explained by general cognitive abilities (although formally trained musicians scored better than non-musicians), as the VAIQ and the Visual Puzzles subtest of the WAIS-IV scores were not significantly correlated. We found significant differences between the scores of the two groups of musicians for the Mini-PROMS, suggesting that formally trained musicians have more aptitude for music than selftaught musicians. If aptitude for music were linked to more vivid auditory imagery, formally trained musicians would have scored higher on the VAIQ than self-taught musicians, but this was not the case. There were, however, weak correlations between scores for the tonal subtests of the Mini-PROMS (Melody and Tuning) and both the VVIQ and VAIQ, while the subtests measuring temporal skills (Accent and Tempo), were not correlated with mental imagery abilities. VAIQ items refer to everyday sounds in relation to their tonal rather than temporal properties, with the exception of two references to "speed." It may be that the tonal and temporal properties of sounds are represented in different ways, with tonal information being easier to imagine than temporal information (Janata & Paroo, 2006). Another possibility is that participants who are asked to 10 Musicae Scientiae 00(0) report the "vividness" of an auditory image focus only on its stable aspects. This is supported by Colley et al.'s (2018) finding that temporal synchronization ability, measured objectively, was predicted by the Control subscale of the BAIS, which assesses how easily an individual can change their mental image of a sound. They also found no significant correlation between temporal synchronization ability and the Vividness subscale of the BAIS. Finally, it can be speculated that the weak correlations between participants' scores for Melody and Tuning subtests of the MiniPROMS and the VVIQ with eyes closed are attributable to a common, higher order brain mechanism. For example, different brain networks are activated when individuals experience different mental states (e.g., when their eyes are open and when their eyes are closed). Interoceptive awareness, often experienced when individuals' eyes are closed, is associated with imagination and recall of sensory information such as auditory information (Xu et al., 2014). The same brain networks may therefore have been activated when participants undertook imagery tasks in the eyes closed condition and the tonal subtests of the Mini-PROMS. Although the VAIQ scores of formally trained and self-taught musicians did not differ significantly, there was a larger effect size for self-taught musicians. It could be that self-taught musicians who do not read music are more reliant than formally trained musicians on mental imagery when they perform; this could in turn be linked to enhanced auditory imagery skills. Next, we might ask why people who play musical instruments score higher on a measure of vividness of auditory imagery than people who do not play a musical instrument. Perhaps this is why people decide to start learning an instrument. Or perhaps playing music enhances auditory imagery ability, as suggested by the small correlations we found between years of music training and/or musical activity and VAIQ scores. It may be that musicians have better auditory imagery because they have a better knowledge than non-musicians of the physical rules that govern the relationships between sounds and their sources (see Gaver, 1993a, 1993b; Spence, 2011 for an overview of cross-modal correspondences). For example, small objects are known to produce high-frequency sounds, whereas large objects are known to produce low-frequency sounds (Grassi, 2005; Grassi et al., 2013). The materials of which objects are made have particular sonic characteristics (Giordano & McAdams, 2006), and the shapes of objects determine the sounds they produce (Kunkler-Peck & Turvey, 2000). Differences between musical instruments make these relationships apparent; for example, compare the size and sound of a violin with those of a double bass, or the sound of a metal vibraphone with that of wooden marimba. Musicians who have been exposed to musical instruments, and who have had the opportunity to associate their physical characteristics with the sounds they make, learn the laws of physics that underlie such associations and can exploit them when asked to imagine everyday sounds. This proposition could be tested by developing a training method for boosting associations between the characteristics of objects and the sounds they make, and finding out if such training also improves auditory imagery. Finally, the VAIQ should be fully validated in future research. Convergent and discriminant validity could be tested, for example, by using the auditory subscales of published questionnaires such as the Plymouth Sensory Imagery Questionnaire (Andrade et al., 2014), and the BAIS; the comparison with the BAIS would also have the advantage of potentially highlighting the differences between auditory imagery for everyday sounds and for musical sounds, as the BAIS is more oriented toward music.

Conclusion The present research assessed self-reported mental imagery abilities in participants with different levels of musical skill. Specifically, we wanted to compare the auditory and visual mental Talamini et al. 11 imagery abilities of formally trained and self-taught musicians and non-musicians, to find out if musical experience is associated with enhanced imagery abilities in general or auditory stimuli in particular. To test this, we compared participants' scores for the VVIQ with their scores for its auditory counterpart, the VAIQ. The findings confirm that music training and/or musical activity are associated with superior mental imagery for everyday auditory stimuli, but not visual stimuli. Future studies should shed light on the origin of this advantage.

Funding The author(s) received no financial support for the research, authorship, and/or publication of this article.

ORCID iD Francesca Talamini https://orcid.org/ 0000-0001-6083-0282

Supplemental material Supplemental material for this article is available online.

Notes 1. Musical aptitude is considered independent of music training, although music training can of course improve this particular trait (Swaminathan & Schellenberg, 2018). 2. https://webapp.uibk.ac.at/psychologie/musiquote/index.php/514829/ lang-en 3. This p value was not significant when the Holm–Bonferroni correction was applied, p=.066.

References Aleman, A., Nieuwenstein, M. R., Böcker, K. B. E., & de Haan, E. H. F. (2000). Music training and mental imagery ability. Neuropsychologia, 38, 1664–1668. Andrade, J., May, J., Deeprose, C., Baugh, S.-J., & Ganis, G. (2014). Assessing vividness of mental imagery: The Plymouth Sensory Imagery Questionnaire. British Journal of Psychology, 105(4), 547–563. Antonietti, A., & Crespi, M. (1995). Analisi di tre questionari per la valutazione della vividezza dell'immagine mentale (Analysis of three questionnaires for assessing the vividness of mental image) [Technical report]. Department of Psychology, Catholic University of the Sacred Heart. Baddeley, A. D., & Andrade, J. (2000). Working memory and the vividness of imagery. Journal of Experimental Psychology: General, 129(1), 126–145. Bailes, F. A. (2006). The use of experience-sampling methods to monitor musical imagery in everyday life. Musicae Scientiae, 10(2), 173–190. Benjamini, Y., & Hochberg, Y. (1995). Controlling the false discovery rate: A practical and powerful approach to multiple testing. Journal of the Royal Statistical Society: Series B (Methodological), 57(1), 289–300. Bishop, L., Bailes, F., & Dean, R. T. (2013a). Musical expertise and the ability to imagine loudness. PLOS ONE, 8(2), Article e56052. Bishop, L., Bailes, F., & Dean, R. T. (2013b). Musical imagery and the planning of dynamics and articulation during performance. Music Perception: An Interdisciplinary Journal, 31(2), 97–117. Bishop, L., Bailes, F., & Dean, R. T. (2014). Performing musical dynamics. Music Perception: An Interdisciplinary Journal, 32(1), 51–66. Brochard, R., Dufour, A., & Després, O. (2004). Effect of musical expertise on visuospatial abilities: Evidence from reaction times and mental imagery. Brain and Cognition, 54(2), 103–109. Brodsky, W., Henik, A., Rubinstein, B.-S., & Zorman, M. (2003). Auditory imagery from musical notation in expert musicians. Perception & Psychophysics, 65(4), 602–612. Brodsky, W., Kessler, Y., Rubinstein, B.-S., Ginsborg, J., & Henik, A. (2008). The mental representation of music notation: Notational audiation. Journal of Experimental Psychology. Human Perception and Performance, 34(2), 427–445. Brown, R. M., & Palmer, C. (2013). Auditory and motor imagery modulate learning in music performance. Frontiers in Human Neuroscience, 7, Article 320. Bywaters, M., Andrade, J., & Turpin, G. (2004). Determinants of the vividness of visual imagery: The effects of delayed recall, stimulus affect and individual differences. Memory, 12(4), 479–488. Campos, A., & Fuentes, L. (2016). Musical studies and the vividness and clarity of auditory imagery. Imagination, Cognition and Personality, 36(1), 75–84. Campos, A., González, M. A., & Amor, A. (2002). The Spanish version of the Vividness of Visual Imagery Questionnaire: Factor structure and internal consistency reliability. Psychological Reports, 90(2), 503–506. Campos, A., & Pérez-Fabello, M. J. (2009). Psychometric quality of a revised version Vividness of Visual Imagery Questionnaire. Perceptual and Motor Skills, 108(3), 798–802. Clark, T., & Williamon, A. (2012). Imagining the music: Methods for assessing musical imagery ability. Psychology of Music, 40(4), 471–493. Colley, I. D., Keller, P. E., & Halpern, A. R. (2018). Working memory and auditory imagery predict sensorimotor synchronisation with expressively timed music. Quarterly Journal of Experimental Psychology, 71(8), 1781–1796. Connolly, C., & Williamon, A. (2004). Mental skills training. In A. Williamon (Ed.), Musical excellence (pp. 221–245). Oxford University Press. Di Nuovo, S. F., & Angelica, A. (2016). Musical skills and perceived vividness of imagery: Differences between musicians and untrained subjects. Annali Della Facoltà Di Scienze Della Formazione Università Degli Studi Di Catania, 14, 3–13. Gaver, W. W. (1993a). How do we hear in the world? Explorations in ecological acoustics. Ecological Psychology, 5(4), 285–313. Gaver, W. W. (1993b). What in the world do we hear? An ecological approach to auditory event perception. Ecological Psychology, 5(1), 1–29. Giordano, B. L., & McAdams, S. (2006). Material identification of real impact sounds: Effects of size variation in steel, glass, wood, and plexiglass plates. The Journal of the Acoustical Society of America, 119(2), 1171–1181. Grassi, M. (2005). Do we hear size or sound? Balls dropped on plates. Perception & Psychophysics, 67(2), 274–284. Grassi, M., Pastore, M., & Lemaitre, G. (2013). Looking at the world with your ears: How do we get the size of an object from its sound? Acta Psychologica, 143(1), 96–104. Gregg, M. J., Clark, T. W., & Hall, C. R. (2008). Seeing the sound: An exploration of the use of mental imagery by classical musicians. Musicae Scientiae, 12(2), 231–247. Guarnera, M., Pellerone, M., Commodari, E., Valenti, G. D., & Buccheri, S. L. (2019). Mental images and school learning: A longitudinal study on children. Frontiers in Psychology, 10, Article 2034. Hall, C. R., & Martin, K. A. (1997). Measuring movement imagery abilities: A revision of the Movement Imagery Questionnaire. Journal of Mental Imagery, 21(1-2), 143–154. Halpern, A. R. (2015). Differences in auditory imagery self-report predict neural and behavioral outcomes. Psychomusicology: Music, Mind, and Brain, 25(1), 37–47. Herholz, S. C., Lappe, C., Knief, A., & Pantev, C. (2008). Neural basis of music imagery and the effect of musical expertise.

European Journal of Neuroscience, 28(11), 2352–2360. Holmes, E. A., Mathews, A., Dalgleish, T., & Mackintosh, B. (2006). Positive interpretation training: Effects of mental imagery versus verbal training on positive mood. Behavior Therapy, 37(3), 237–247. Holmes, P. (2005). Imagination in practice: A study of the integrated roles of interpretation, imagery and technique in the learning and memorisation processes of two experienced solo performers. British Journal of Music Education, 22(3), 217–235. Isaac, A., Marks, D. F., & Russell, D. G. (1986). An instrument for assessing imagery of movement: The Vividness of Movement Imagery Questionnaire (VMIQ). Journal of Mental Imagery, 4, 23–30. Talamini et al. 13 Isaac, A. R., & Marks, D. F. (1994). Individual differences in mental imagery experience: Developmental changes and specialization. British Journal of Psychology London, England, 85(Pt. 4), 479–500. Janata, P., & Paroo, K. (2006). Acuity of auditory images in pitch and time. Perception & Psychophysics, 68(5), 829–844. Keller, P. E. (2012). Mental imagery in music performance: Underlying mechanisms and potential benefits. Annals of the New York Academy of Sciences, 1252(1), 206–213. Keller, P. E., & Appel, M. (2010). Individual differences, auditory imagery, and the coordination of body movements and sounds in musical ensembles. Music Perception, 28(1), 27–46. Keller, P. E., & Koch, I. (2008). Action planning in sequential skills: Relations to music performance. Quarterly Journal of Experimental Psychology, 61(2), 275–291. Kosslyn, S. M., Ganis, G., & Thompson, W. L. (2001). Neural foundations of imagery. Nature Reviews Neuroscience, 2(9), 635–642. Kunkler-Peck, A. J., & Turvey, M. T. (2000). Hearing shape. Journal of Experimental Psychology: Human Perception and Performance, 26(1), 279–294. Liu, L. L., & Park, D. C. (2004). Aging and medical adherence: The use of automatic processes to achieve effortful things. Psychology and Aging, 19(2), 318–325. Marks, D. F. (1973). Visual imagery differences in the recall of pictures. British Journal of Psychology, 64(1), 17–24. Marsh, H. W., Hau, K. T., Balla, J. R., & Grayson, D. (1998). Is more ever too much? The number of indicators per factor in confirmatory factor analysis. Multivariate Behavioral Research, 33(2), 181–220. McAvinue, L. P., & Robertson, I. H. (2007). Measuring visual imagery ability: A review. Imagination, Cognition and Personality, 26(3), 191–211. Mizuguchi, N., Nakata, H., Uchida, Y., & Kanosue, K. (2012). Motor imagery and sport performance. The Journal of Physical Fitness and Sports Medicine, 1(1), 103–111. Moulton, S. T., & Kosslyn, S. M. (2009). Imagining predictions: Mental imagery as mental emulation. Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences, 364(1521), 1273–1280. Münte, T. F., Altenmüller, E., & Jäncke, L. (2002). The musician's brain as a model of neuroplasticity. Nature Reviews Neuroscience, 3(6), 473–478. Nelis, S., Holmes, E. A., Griffith, J. W., & Raes, F. (2014). Mental imagery during daily life: Psychometric evaluation of the Spontaneous Use of Imagery Scale (SUIS). Psychologica Belgica, 54(1), 19–32. Palmiero, M., Cardi, V., & Belardinelli, M. O. (2011). The role of vividness of visual mental imagery on different dimensions of creativity. Creativity Research Journal, 23(4), 372–375. Palmiero, M., Guariglia, P., Crivello, R., & Piccardi, L. (2020). The relationships between musical expertise and divergent thinking. Acta Psychologica, 203, Article 102990. Pecenka, N., & Keller, P. E. (2009, August). The relationship between auditory imagery and musical synchronisation abilities in musicians [Conference session]. 7th Triennial Conference of European Society for the Cognitive Sciences of Music (ESCOM), Jyväskylä, Finland. Pfordresher, P. Q., & Halpern, A. R. (2013). Auditory imagery and the poor-pitch singer. Psychonomic Bulletin & Review, 20(4), 747–753. Renner, F., Murphy, F. C., Ji, J. L., Manly, T., & Holmes, E. A. (2019). Mental imagery as a "motivational amplifier" to promote activities. Behaviour Research and Therapy, 114, 51–59. Schifferstein, H. N. J. (2009). Comparing mental imagery across the sensory modalities. Imagination, Cognition and Personality, 28(4), 371–388. Schlaug, G. (2001). The brain of musicians: A model for functional and structural adaptation. Annals of the New York

Academy of Sciences, 930(1), 281–299. http://muziekenzorg.nl/userfiles/files/kennisbank/neuroscience/the%20brain%20of%20musicians%23.pdf Sergent, J. (1993). Music, the brain and Ravel. Trends in Neurosciences, 16(5), 168–172. Shaw, G. A., & DeMers, S. T. (1986). The relationship of imagery to originality, flexibility and fluency in creative thinking. Journal of Mental Imagery, 10(1), 65–74. Sheehan, P. S. (1967). A shortened form of Betts' questionnaire upon mental imagery. Journal of Clinical Psychology, 23(3), 386–389. 14 Musicae Scientiae 00(0) Spence, C. (2011). Crossmodal correspondences: A tutorial review. Attention, Perception, & Psychophysics, 73(4), 971–995. Swaminathan, S., & Schellenberg, E. G. (2018). Musical competence is predicted by music training, cognitive abilities, and personality. Scientific Reports, 8(1), 1–7. Swaminathan, S., & Schellenberg, E. G. (2019). Music training and cognitive abilities: Associations, causes, and consequences. In M. H. Thaut, D. A. Hodges, S. Swaminathan, & E. G. Schellenberg (Eds.), The Oxford handbook of music and the brain (pp. 644–670). Oxford University Press. Vasquez, N. A., & Buehler, R. (2007). Seeing future success: Does imagery perspective influence achievement motivation? Personality & Social Psychology Bulletin, 33(10), 1392–1405. Wechsler, D. (2008). Wechsler Adult Intelligence Scale–Fourth Edition. Pearson. Wong, S. S. H., & Lim, S. W. H. (2017). Mental imagery boosts music compositional creativity. PLOS ONE, 12(3), Article e0174009. Xu, P., Huang, R., Wang, J., Van Dam, N. T., Xie, T., Dong, Z., . . .Luo, Y. J. (2014). Different topological organization of human brain functional networks with eyes open versus eyes closed. NeuroImage, 90, 246–255. Zendel, B. R., & Alexander, E. J. (2020). Autodidacticism and music: Do self-taught musicians exhibit the same auditory processing advantages as formally trained musicians? Frontiers in Neuroscience, 14, Article 752. Zentner, M., & Strauss, H. (2017). Assessing musical ability quickly and objectively: Development and validation of the Short-PROMS and the Mini-PROMS. Annals of the New York Academy of Sciences, 1400(1), 33–45.