The Thickness of Musical Pitch: Psychophysical evidence for the Whorfian hypothesis

Sarah Dolscheid[1,2] (sarah.dolscheid@mpi.nl) **Shakila Shayan[1]** (shakila.shayan@mpi.nl) **Asifa Majid[1,3]** (asifa.majid@mpi.nl) **Daniel Casasanto[1,3,4]** (casasand@newschool.edu)

[1]Max Planck Institute for Psycholinguistics, Nijmegen, NL
[2]International Max Planck Research School for Language Sciences, Nijmegen, NL
[3]Donders Center for Brain, Cognition, and Behaviour, Radboud University, Nijmegen, NL
[4]Department of Psychology, The New School for Social Research, New York, USA

Abstract

Do the languages that people speak affect the way they think about musical pitch? Here we compared pitch representations in native speakers of Dutch and Farsi. Dutch speakers describe pitches as 'high' (*hoog*) and 'low' (*laag*), but Farsi speakers describe high-frequency pitches as 'thin' (*naazok*) and low-frequency pitches as 'thick' (*koloft*). Differences in language were reflected in differences in performance on two psychophysical pitch reproduction tasks. This was true even though the tasks used entirely nonlinguistic stimuli and responses. To test whether experience using language changes pitch representations, we trained native Dutch speakers to use Farsi-like metaphors, describing pitch relationships in terms of thickness. After training, Dutch speakers' performance on a nonlinguistic psychophysical task resembled native Farsi speakers'. People who use different space-pitch metaphors in language also think about pitch differently. Beyond correlation, language plays a causal role in shaping mental representations of musical pitch.

Keywords: Metaphor; Musical pitch; Psychophysics; Space; Whorfian hypothesis.

Introduction

Speakers often use spatial metaphors to talk about musical pitch. In English, pitches can be *high* or *low*, melodic contours can *rise* or *fall* and people can sing at the *top* or the *bottom* of their range. Are these spatial metaphors merely linguistic conventions, or do they reflect something fundamental about the way people mentally represent musical pitch?

There are several reasons to believe that pitch and space are importantly related in the brain and mind. Amusic patients, who have difficulty discriminating pitch changes, also have deficits in spatial tasks like mental rotation compared to control groups of musicians and non-musicians (Douglas & Bilkey, 2007). Behavioral experiments also demonstrate a systematic relationship between pitch and space in normal participants. In stimulus-response compatibility tasks, participants are faster to press higher response keys to identify high-frequency pitches than to press lower response keys, and vice versa for low-frequency pitches (Lidji, Kolinsky, Lochy, & Morais, 2007; Rusconi, Kwan, Giordano, Umiltà, & Butterworth, 2006). Beyond binary high-low correspondences, psychophysical pitch reproduction tasks show that pitch maps onto vertical space in a continuous analog fashion (Casasanto, 2010).

Pitch and vertical space have been found to interact even in prelinguistic infants. In a preferential looking task, 3- to 4-month-olds preferred congruent trials (in which visuospatial height and pitch height corresponded) over incongruent trials (Walker et al., 2010; see also Wagner, et al., 1981). There is, thus, converging evidence that people mentally represent pitch in terms of vertical space, just like they talk about it.

Crucially, however, not everybody talks about pitch the same way. In spite of the apparent predominance of the 'high-low' metaphor, some languages do not metaphorize pitch spatially. The Kpelle people of Liberia for instance, talk about *high* and *low* pitches as *light* and *heavy*. The Suyá people of the Amazon basin call high pitches *young* and low pitches *old*, and the Bashi people of central Africa call high pitches *weak* and low pitches *strong* (Eitan & Timmers, 2010).

Even languages that use spatial metaphors for pitch may not use the same vertical metaphors that are familiar to English speakers. For the Manza of Central Africa, high pitches are *small* and low pitches *large* (Stone, 1981). In other languages like Farsi, Turkish and Zapotec (spoken in the Sierra Sur of Mexico) high pitches are *thin* and low pitches *thick* (Shayan, Ozturk & Sicoli, 2011).

This linguistic diversity in pitch vocabulary raises a question: Do people who use different metaphors in language mentally represent pitch differently? If so, how deep are the effects of language on musical pitch? Could language shape the nonlinguistic representations that people use for perceiving or producing musical pitches, even when they are not using language?

The question of linguistic relativity, often associated with the writings of Benjamin Whorf (1956), has been extensively debated in domains like "time" (e.g., Boroditsky, 2001; Casasanto 2008), "space" (e.g., Majid, Bowerman, Kita, Haun, & Levinson, 2004), "motion" (e.g., Gennari, Sloman, Malt & Fitch, 2002; Papafragou, Hubert, & Trueswell, 2008), and "color" (e.g., Regier & Kay, 2009), but little is known about effects of language on pitch representation.

A first hint that people who use different pitch metaphors conceptualize pitch differently comes from co-speech gestures. Consistent with the Manzas' linguistic coding of pitches as *small* and *large*, speakers have been observed continuously lowering their hand in space while referring to the *smaller* (i.e., higher) pitches, contrary to the English high-low mapping (Ashley, 2004). This suggests that people may conceptualize pitch consistent with their pitch vocabulary. However, on a skeptical interpretation of these data, gestures that match the co-occurring speech may only reveal conventions for communicating about musical pitches, not modes of conceptualizing them. Alternatively, they may reveal a 'shallow' influence of language on thought, indicating that people do indeed conceptualize pitch in language-specific ways, but only while they are packaging their thoughts into words (i.e., while they are 'thinking for speaking', see Slobin, 1996).

Further evidence for cross-linguistic differences in pitch representation comes from a developmental study in Farsi-, Turkish-, and German-speaking children (Shayan, et al., in preparation). Children were asked to match tones of different pitches with toy snakes of different thicknesses. Turkish- and Farsi-speaking children reliably matched the low-pitched sounds to the thicker snake and the high-pitched sounds to the thinner snake, consistent with thick-thin metaphors in their languages. German children, who are exposed to high-low metaphors in their language, did not show the same thick-thin response pattern.

A persistent challenge in testing relationships between language and nonlinguistic mental representations is devising truly nonlinguistic tasks. Here we tested pitch representations in speakers of one language that uses 'height' metaphors (Dutch) and another that uses 'thickness' metaphors (Farsi), using a pair of psychophysical tasks with non-linguistic stimuli and responses (adapted from Casasanto 2010). In one task (Height Interference), participants saw lines of varying heights while listening to tones of different pitches. After each tone, participants reproduced the pitch by singing it back. In the other task (Thickness Interference) participants saw lines varying in thickness while hearing tones of different pitches, and sang back the pitches that they had heard.

In both tasks, the spatial information was irrelevant, and spatial variation was orthogonal to variation in pitch. As such, the spatial dimension of the stimuli served as a distractor: a piece of information that could potentially interfere with performance on the pitch reproduction task. We reasoned that if Dutch and Farsi speakers' concepts of pitch were similar irrespective of the languages they speak, then performance on these tasks should not differ between language groups. On the other hand, if Dutch and Farsi speakers mentally represent pitch the way they talk about it, using different kinds of spatial representations, they should show contrasting patterns of cross-dimensional interference: Dutch speakers' pitch estimates should be more strongly affected by irrelevant height information, and Farsi speakers' by irrelevant thickness information.

Experiment 1:

Do people think about pitch like they talk about it?

Experiment 1 tested whether the relationships between space and pitch found in Dutch and Farsi speakers' linguistic metaphors are also present in their nonlinguistic pitch representations.

Methods

Participants Native Dutch speakers (N=40) and native Farsi speakers (N=40) participated in this study for payment. Half of the participants from each language group performed the Height Interference task (N=20 Dutch, 20 Farsi) and half performed the Thickness Interference task (N=20 Dutch, 20 Farsi). One additional Farsi-speaking participant was tested, but was excluded for performing the task incorrectly. Dutch participants were recruited from the Max Planck Institute participant pool. Farsi speakers were recruited from Nijmegen and Delft.

Materials For the Height Interference experiment, horizontal lines intersected a vertical reference line at one of nine different locations (ranging from 80 to 720 pixels from bottom to top of the computer screen, in 80 pixel increments). For the Thickness Interference experiment, a vertical line appeared in the middle of the screen in one of nine thicknesses (ranging from 8 to 72 pixels in 8 pixel increments). Variation in thickness was thus proportional to variation in height. In each experiment, the nine different lines were fully crossed with nine different pitches ranging from C4 to G#4 in semitone increments, to produce 81 distinct trials. All stimuli were presented on a pc laptop (screen resolution = 1024x768 pixels) using Presentation software (www.neurobs.com). Lines were presented in white on a grey background (350 pixels wide) which was surrounded by black frames. Auditory stimuli were created using Audacity software (http://audacity.sourceforge.net/) and presented through sealed headphones.

Singing responses were recorded by an EDIROL R-09 recording device, and analyzed using Praat software (http://www.fon.hum.uva.nl/praat/) by a coder blind to the corresponding spatial stimuli. The approximate temporal midpoint of each response was determined by visual inspection of the waveform. The average fundamental frequency (F_0) of each sung response was extracted from an interval spanning 300 ms before and after the estimated temporal midpoint, to ensure that measured F_0 was representative of the whole response.

Instructions were translated by native speakers of Dutch and Farsi, and contained no space-pitch metaphors. Although the language of instructions differed across language groups, the tasks themselves comprised only nonlinguistic stimuli (lines and tones) and responses (sung tones).

Procedure Participants were asked to watch the lines and listen to the pitches carefully, and to sing back the pitches as accurately as possible. They were tested individually, and received written instructions prior to the start of the experiment in their native language.

After three practice trials, participants were presented each of the 81 line-pitch pairings one at a time, in random order. Line-pitch stimuli were presented for 2 seconds each. Immediately after each stimulus, a picture of a microphone appeared in the center of the screen indicating that the participants had 2 seconds to sing back the pitch they had heard. Each response period was followed by an inter-trial interval of 500 milliseconds. After 40 trials, participants were given a break, the duration of which was self-paced. Testing lasted about 15 minutes, and was followed by a debriefing.

Results

Pitch estimation, cross-domain effects

The effects of irrelevant spatial information on pitch reproduction were first tested for each group and each task, individually. The values of the height and thickness stimuli were normalized. For each participant we computed the normalized slope of the effect of the height or thickness of the stimuli on participants' reproduced pitches (figure 1 presents data averaged over participants). In Dutch speakers, the spatial height of the stimuli influenced pitch estimates as predicted by 'height' metaphors in Dutch ($t(19)=2.70$, $p=.01$), but the thickness of stimuli had no significant effect on pitch reproduction ($t(19)=0.57$, *ns*). Farsi speakers showed the opposite pattern: thickness influenced pitch estimates as predicted by 'thickness' metaphors in Farsi ($t(19)=2.09$, $p=.05$), but height had no significant effect on pitch reproduction ($t(19)=1.16$, *ns*).

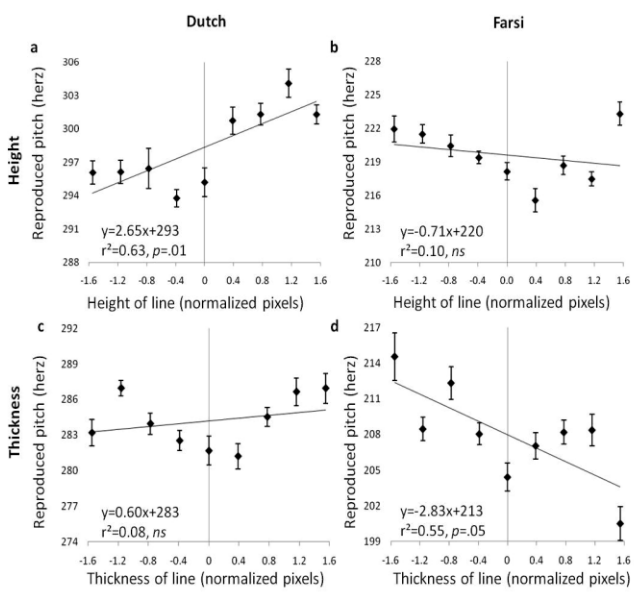

Figure 1: Results of Experiment 1. Effects of height interference (top) and thickness interference (bottom) on pitch estimates in speakers of Dutch (left) and Farsi (right). Error bars indicate s.e.m.

To test for the predicted interaction of Language (Dutch, Farsi) and Task (Height Interference, Thickness Interference), the normalized slopes from the thickness task were multiplied by -1. This multiplication was necessary because, interestingly, the relationship between spatial magnitude and frequency reverses between height and thickness metaphors. Greater spatial height corresponds to higher frequency, but greater spatial thickness corresponds to *lower* frequency. Multiplying the slopes by -1 for one of the tasks aligns the space and pitch continuums (i.e., the slope then indicates the same relationship between spatial magnitude and frequency for both tasks). The normalized slopes of the effects of space on pitch reproduction were submitted to a 2 x 2 ANOVA. Language interacted with Task to predict the effect of space on pitch estimates ($F(1,79)=10.73$, $p=.002$), consistent with the use of 'height' metaphors in Dutch and 'thickness' metaphors in Farsi (figure 2.) There were no main effects.

In planned pairwise comparisons, the effect of height interference was greater in Dutch speakers than in Farsi speakers ($t(38)=2.65$, $p=.01$) and conversely, the effect of thickness interference was greater in Farsi speakers than in Dutch speakers ($t(38)=2.00$, $p=.05$). In Dutch speakers the effect of height interference was greater than the effect of thickness interference ($t(38)=2.26$, $p=.03$, and in Farsi speakers the effect of thickness interference was greater than the effect of height interference ($t(38)=2.38$, $p=.02$, all tests two-tailed).

In both Dutch and Farsi culture, higher pitches are written higher on the musical staff. In principle, differences in experience using this nonlinguistic cultural convention could be responsible for the observed effects, as opposed to differences in experience using language. A further analysis ruled out this possibility. During debriefing, participants rated how well they read music on a scale of 1-7. When this rating was added as a covariate to the 2 x 2 ANOVA, the interaction between Language and Task remained highly significant, even when differences in music reading ability were controlled ($F(1,79)=10.83$, $p=.002$).

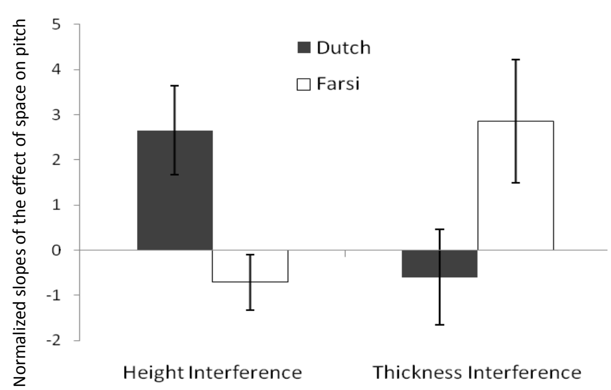

Figure 2: Cross-dimensional interference effects in Experiment 1. Error bars indicate s.e.m.

Pitch estimation, within-domain effects

Further analyses were conducted to ensure that differences in cross-dimensional interference were not due to differences in the accuracy with which participants reproduced pitches. For each participant we computed the slope of the effect of the actual pitches on participants' reproduced pitches (Dutch Height: Average slope=1.02, $t(19)=12.50$, $p=.0001$; Dutch Thickness: Average slope=0.87, $t(19)=12.68$, $p=.0001$; Farsi Height: Average slope=0.61, $t(19)=6.67$, $p=.0001$; Farsi Thickness: Average slope=0.58, $t(19)=6.21$, $p=.0001$). According to a 2 x 2 ANOVA, Language did not interact with Task to predict the effect of actual pitch on estimated pitch ($F(1,79)=0.45$, *ns*). Overall, Dutch speakers' pitch estimates were more accurate than Farsi speakers' ($F(1,79)=16.99$, $p=.0001$), but this main effect of Language on within-domain performance cannot explain the critical Language x Task interaction we found in the cross-domain analysis.

Finally, we conducted a 3-way ANOVA on the normalized slopes, combining the cross-domain and within-domain analyses. There was a 3-way interaction of Language (Dutch, Farsi), Task (Height Interference, Thickness Interference) and Domain (Within-domain effects, Cross-domain effects; $F(1,159)=6.12$, $p=.02$), indicating that the predicted cross-dimensional interference effects were not due to unpredicted differences in within-domain performance.

In summary, performance on these nonlinguistic tasks suggests that people who use different metaphors in their native languages form correspondingly different mental representations of musical pitch.

Experiment 2:

Does language shape pitch representations?

Did language give rise to the cross-linguistic differences in performance reported for the Height and Thickness interference experiments? Although the data from Experiment 1 closely follow predictions based on language, they are nevertheless correlational. A 2-part training study was conducted to investigate a causal role for language.

Dutch speakers were trained to complete sentences about pitch relationships using Farsi-like thickness metaphors (Thickness Training), or using the familiar high-low metaphors (Height Training) as a control. To determine whether this linguistic training had an effect on nonlinguistic pitch representations, we then tested all participants on the Thickness Interference task described in Experiment 1. If experience using the pitch metaphors in their native language causes Farsi speakers to think about pitch using mental representations of spatial thickness, then repeatedly using thickness metaphors during training should transiently strengthen Dutch speakers' nonlinguistic thickness-pitch mappings, and should increase the effect of irrelevant thickness information on pitch estimation.

Methods

Participants Native Dutch speakers (N=60) participated for payment. Half were assigned to the Thickness Training task, and the other half to the Height Training task.

Materials Participants completed 196 fill-in-the-blank sentences using the words *dunner* (thinner) and *dikker* (thicker) in the Thickness Training condition and the words *hoger* (higher) or *lager* (lower) in the Height Training condition. In both tasks, half of the sentences compared the length or thickness of physical objects (e.g., A tower is higher / lower than a blade of grass; A pillar is thicker / thinner than a finger); the other half compared the pitches of different sounds (e.g., A flute sounds higher / lower than a bass; A flute sounds thicker / thinner than a bass). Stimuli were presented on a pc laptop using Presentation software.

Procedure Participants saw 3 correctly-completed example sentences before the experiment started. They were then presented one sentence at a time and instructed to fill in the blank by typing the correct response. They were not told whether 'thicker' or 'thinner' meant 'higher' or 'lower'; rather, they were left to infer the correct mapping based on the example sentences, and on the written feedback they received after each trial (either 'goed' (correct) for correct responses or 'fout' (incorrect) for incorrect responses). Training took about 20 minutes. After the training phase, all participants performed the Thickness Interference task used in Experiment 1.

Results and Discussion

Training Phase Participants filled in the blanks with high accuracy for both the Height Training (Mean %correct=0.99, SD=0.01) and the Thickness Training task (Mean %correct=0.99, SD=0.01). Accuracy did not differ between tasks (difference of means=0.002, $t(58)=.34$, *ns*).

Test Phase In the Thickness Interference task, the effect of thickness on pitch estimation was highly significant following Thickness Training (Slope=1.46, $p=.003$), but not following Height Training (Slope=0.08, *ns*; difference of slopes=1.38, $t(58)=1.84$, $p=.07$, two-tailed). The effect of thickness on pitch estimation in thickness-trained participants was statistically indistinguishable from the effect in native Farsi speakers (difference of slopes=1.39, $t(48)=1.12$, *ns*), and was significantly greater than the effect in untrained Dutch speakers (difference of slopes=2.06, $t(48)=2.02$, $p=.05$).

Experience using language can change non-linguistic mental representations of musical pitch.

General Discussion

Dutch and Farsi speakers, who use different metaphors for pitch in language, also form correspondingly different nonlinguistic pitch representations. We show this via a double-dissociation between Dutch- and Farsi-speakers' performance on a pair of nonlinguistic psychophysical tasks.

Dutch speakers, who talk about pitches as 'high' and 'low', incorporated irrelevant height information into their pitch estimates (but ignored irrelevant thickness information). Farsi speakers, who talk about pitches as 'thin' and 'thick', incorporated irrelevant thickness information into their pitch estimates (but ignored irrelevant height information). When Dutch speakers were trained to use Farsi-like metaphors, they showed the same pattern of cross-dimensional thickness interference as native Farsi speakers. Beyond demonstrating a language-thought correlation, results show that metaphors in language can play a causal role in shaping nonlinguistic mental representations of musical pitch.

Beyond Thinking for Speaking

On one influential view of the relationship between language and thought, patterns in language can influence nonlinguistic mental representations, but only (or primarily) while people are packaging their thoughts into words (Slobin, 1987, 1996), or while they are performing tasks for which verbal codes can be helpful (e.g., Gennari, Sloman, Malt & Fitch, 2002; Papafragou, Hubert, & Trueswell, 2008). But these 'shallow' effects of language on high-level language-mediated thinking are only one sort of linguistic relativity effect. The present results support the proposal that language can also have 'deep' effects on people's low-level perceptuo-motor abilities (Casasanto, 2008), such as their ability to perceive and reproduce musical pitches.

Although participants were not producing or comprehending language during these pitch reproduction tasks, it remains possible that they were using language covertly to label the stimuli. However, this is unlikely to account for the observed pattern of cross-dimensional interference, for a combination of reasons. First, the increments of space and pitch were too fine-grained to be labeled using ordinary (non-technical) words: e.g., relative to the other pitches, C4 and C#4 would both be labeled 'low'. Yet, participants could perceive and reproduce values along the analog space and frequency continuums that could not be discriminated using their lexical categories. (Relative coding like *higher than the last* would be of little use since stimulus order was random.)

Most importantly, covertly labeling the *pitches* using spatial words could not possibly result in the observed pattern of spatial interference because variation in pitch was orthogonal to variation in space. Only verbally labeling the *spatial* height or thickness of the stimuli could, in principle, contribute to the spatial interference effects. Yet, labeling the spatial dimension of stimuli would be unmotivated given that this information was always task-irrelevant, and it would be unhelpful given that (a) labeling the irrelevant dimension would interfere with labeling the relevant one, and (b) the irrelevant spatial information was often in conflict with the relevant pitch information (e.g., a 'high' line would often appear during a 'low' pitch).

Rather than an online effect of using verbal labels, we propose that the observed cross-dimensional interference resulted from analog relationships between space and pitch in long-term memory, which are partly conditioned by language. Suppose each time people produce or understand a spatial metaphor for pitch they activate the corresponding mental metaphor: an associative mapping between nonlinguistic mental representations in the source domain (space) and target domain (pitch). Over time, speakers of a 'height' language like Dutch would strengthen the height-pitch mapping at the expense of any competitors, such as the thickness-pitch mapping -- and vice versa for speakers of a 'thickness' language like Farsi. This associative learning model is supported by the training effect we report in Experiment 2.

Malleability of mental metaphors

The finding that Dutch speakers' mental representations of pitch could be retrained to resemble Farsi speakers' in only 20 minutes may seem surprising, but such rapid retraining effects are not unprecedented (see Boroditsky, 2001; Casasanto & Bottini, 2010; Casasanto & Chrysikou, 2011; Fischer, Mills, & Shaki, 2010). In particular, the present results echo a relationship between space-time metaphors in language and in thought. Many languages like English tend to use distance metaphors to talk about duration (e.g., *a long time*). In other languages like Greek, duration tends to be metaphorized as an amount of a substance in 3-dimensional space (e.g., *poli ora*, tr. 'much time'). In a pair of nonlinguistic psychophysical tasks analogous to those presented here, English and Greek speakers were asked to reproduce the duration of stimuli while ignoring their spatial extent (in one task) or their volume (in the other). Irrelevant distance information influenced English speakers' time estimates more strongly than irrelevant volume information, but the opposite pattern was found in Greek speakers: Mental metaphors for time mirrored linguistic metaphors. After about 20 minutes of exposure to Greek-like space-time metaphors in language, however, the effect of volume on time reproduction was just as strong in 'volume-trained' English speakers as in native Greek speakers (Casasanto, 2008).

The effect of short-term training with new linguistic metaphors is presumably temporary, but the same associative learning mechanisms that change people's performance in the laboratory may also influence mental metaphors in the course of ordinary language use. Using one's native language may serve as a natural 'training task', encouraging the habitual use of some nonlinguistic metaphorical mappings more than others (for a discussion of how brief training can transiently overwhelm long-term training see Casasanto & Bottini, 2010.)

Origins of space-pitch mappings

Does language establish mental metaphors between space and pitch in the first place, or does it modify preexisting mappings? Given that space-pitch mappings have been demonstrated in prelinguistic infants (e.g., Walker et al., 2010), the latter seems more likely. In principle, both height-pitch and thickness-pitch mappings could be present in children's minds; the relative strength of these mappings could be adjusted subsequently, according to the relative frequencies of these metaphors in the languages they acquire.

This proposal raises the question: Where do these mental metaphors come from, if not from language? Both of the space-pitch metaphors we explore here could plausibly be based on correspondences in the physical world. The relationship between thickness and pitch is evident in musical instruments (e.g., thicker strings produce lower tones). As people produce higher pitches the larynx rises; as they produce lower pitches it descends.

Yet, these 'just so stories' about the physical origins of mental metaphors should be interpreted with caution (or tested directly, e.g., Casasanto & Chrysikou, 2011). It is easy to find other physical regularities that predict different relationships between pitch and space (e.g., taller people, tend to have *lower* voices). Furthermore, it remains an open question to what extent space-pitch mappings in our minds emerge in developmental time, as individuals track experiential regularities (e.g., Casasanto, 2009; Lakoff & Johnson, 1999) or in evolutionary time, as the neural substrates of spatial cognition were exapted for non-spatial functions.

Metaphoric representation of concrete experiences

The great majority of psychological experiments on metaphor have tested for mappings from concrete domains like *space* to abstract domains that can never be perceived through the senses, such as *time* (e.g., Boroditsky, 2001), *intimacy* (Williams & Bargh, 2008), and *similarity* (Casasanto, 2009). Unlike these target domains, however, musical pitch can be perceived directly: Why should pitch be represented metaphorically, in language or thought? Although pitch is 'concrete' in the sense that it is perceptible, arguably it can only be perceived via one sensory channel: audition. Compared with the metaphoric source domain of space, which can be perceived multimodally (e.g., spatial distance can be judged based on sight, sound, touch, or even smell), pitch is *relatively* abstract. Here we show that mental metaphors can structure even target domains that are, themselves, grounded in perception, and that experience with language can shape these metaphorical mappings.

References

Ashley, R. (2004). Musical pitch space across modalities: Spatial and other mappings through language and culture. In P. W. S. Lipscomb, R. Ashley, R. Gjerdingen (Ed.), *Proceedings of the 8th International Conference on Music Perception and Cognition*. Adelaide: Causal Productions.

Boroditsky, L. (2001). Does language shape thought? Mandarin and English speakers' conceptions of time. *Cognitive Psychology*, 43(1), 1-22.

Casasanto, D. (2008). Who is afraid of the big bad Whorf? Cross-linguistic differences in temporal language and thought. *Language Learning, 58(1),* 63-79.

Casasanto, D. (2009). When is a Linguistic Metaphor a Conceptual Metaphor? In V. Evans & S. Pourcel (Eds.), *New Directions in Cognitive Linguistics* (pp. 127-145). Amsterdam: John Benjamins.

Casasanto, D. (2010). Space for Thinking. In V. Evans &. P. Chilton (Eds.), *Language, Cognition and Space: The State of the Art and New Directions* (pp. 453-478). London: Equinox Publishing.

Casasanto, D., Bottini, R. (2010). Can Mirror-Reading Reverse the Flow of Time? In N.S. Hölscher, T.F. Shipley, M. Olivetti Belardinelli, J.A. Bateman (Eds.), *Spatial Cognition VII* (pp. 335-345). Berlin Heidelberg: Springer.

Casasanto, D., & Chrysikou, E. G. (2011). When Left it 'Right': Motor fluency shapes abstract concepts. *Psychological Science*, 22(4), 419-22

Douglas, K. M., & Bilkey, D. K. (2007). Amusia is associated with deficits in spatial processing. *Nat. Neurosci.*, 10(7), 915-21.

Eitan, Z., & Timmers, R. (2010). Beethoven's last piano sonata and those who follow crocodiles: cross-domain mappings of auditory pitch in a musical context. *Cognition*, 114(3), 405-22.

Fischer, M. H., Mills, R. A, & Shaki, S. (2010). How to cook a SNARC: number placement in text rapidly changes spatial-numerical associations. *Brain and Cognition*, 72(3), 333-6.

Gennari, S. P., Sloman, S., Malt, B. C., & Fitch, W. T. (2002). Motion events in language and cognition. *Cognition*, 83, 49-79.

Lakoff J., Johnson, M. (1999). *Philosophy in the Flesh*. Basic Books.

Lidji, P., Kolinsky, R., Lochy, A., & Morais, J. (2007). Spatial associations for musical stimuli: a piano in the head? *JEP: HPP*, 33(5), 1189-207.

Majid, A., Bowerman, M., Kita, S., Haun, D. B. M., & Levinson, S. C. (2004). Can language restructure cognition? The case for space. *Trends in Cognitive Sciences*, 8(3), 108-14.

Papafragou, A., Hulbert, J., & Trueswell, J. (2008). Does language guide event perception? Evidence from eye movements. *Cognition*, 108(1), 155-84.

Regier, T., & Kay, P. (2009). Language, thought, and color: Whorf was half right. *Trends in Cognitive Sciences*, 13(10), 439-46.

Rusconi, E., Kwan, B., Giordano, B. L., Umiltà, C., & Butterworth, B. (2006). Spatial representation of pitch height: the SMARC effect. *Cognition*, 99(2), 113-29.

Shayan, S., Ozturk O., & Sicoli, M. (2011). Thickness of pitch, cross-modal metaphors in Farsi, Turkish and Zapotec, *The Senses and Society*. In press.

Slobin, D. I. (1996). From "thought and language" to "thinking for speaking." In J. J. Gumperz & S. C. Levinson (Eds.), *Rethinking linguistic relativity* (pp. 70–96). Cambridge: Cambridge University Press.

Stone, R. M. (1981). Toward a Kpelle conceptualization of music performance. *Journal of African Folklore*, 94, 188-206.

Wagner, S., Winner, E., Cicchetti, D., & Gardner, H. (1981). "Metaphorical" Mapping in Human Infants. *Child Development*, 52(2), 728.

Walker, P., et al. (2010). Preverbal infants' sensitivity to synaesthetic cross-modality correspondences. *Psychological Science*, 21(1), 21-5.

Whorf, B. L. (1956). In J. B. Carrol (Ed.), *Language, thought, and reality: Selected writings of Benjamin Lee Whorf*. Cambridge: MIT Press.

Williams, L. E., & Bargh, J. A. (2008). Keeping one's distance: The influence of spatial distance cues on affect and evaluation. Psychological Science, 19, 302–308.