



SHORT REPORT

Spatial metaphor in language can promote the development of cross-modal mappings in children

Shakila Shayan,¹ Ozge Ozturk,¹ Melissa Bowerman^{1*} and Asifa Majid^{1,2,3}

1. Max Planck Institute for Psycholinguistics, Nijmegen, The Netherlands

2. Centre for Language Studies, Radboud University, The Netherlands

3. Donders Institute for Brain, Cognition, and Behaviour, Radboud University, The Netherlands

Abstract

Pitch is often described metaphorically: for example, Farsi and Turkish speakers use a ‘thickness’ metaphor (low sounds are ‘thick’ and high sounds are ‘thin’), while German and English speakers use a height metaphor (‘low’, ‘high’). This study examines how child and adult speakers of Farsi, Turkish, and German map pitch and thickness using a cross-modal association task. All groups, except for German children, performed significantly better than chance. German-speaking adults’ success suggests the pitch-to-thickness association can be learned by experience. But the fact that German children were at chance indicates that this learning takes time. Intriguingly, Farsi and Turkish children’s performance suggests that learning cross-modal associations can be boosted through experience with consistent metaphorical mappings in the input language.

Research highlights

- Pitch is often described metaphorically across languages.
- We examined the development of the pitch–thickness mapping cross-linguistically.
- Results suggest that the pitch–thickness association is learned.
- Learning happens faster if the language you speak promotes the mapping.

Introduction

In some people hearing a sound elicits a visual image, for example, a loud sound triggers a bright image while a quiet sound elicits a dim one, or a high pitch suggests a small object while a low pitch invokes a large one (Marks, 1974, 1978, 1987). Synaesthetes are a special population of people who experience these sorts of cross-modal correspondences with exceptional vividness.

Activation of a primary modality (e.g. sound) produces a secondary percept in a non-related modality (e.g. vision). But even in the normal population, cross-modal correspondences are ubiquitous. Experiments with adults and young children without synaesthesia demonstrate consistent patterns of association, such as matching high-pitched sounds to bright objects (e.g. Marks, Hammeal & Bornstein, 1987; Mondloch & Maurer, 2004). Even infants as young as 3 months are sensitive to some of these correspondences (e.g. Walker, Brenner, Spring, Maststock, Slater & Johnson, 2010), which suggests that at least some cross-modal mappings have a built-in neurological basis that is robust over developmental time (Marks *et al.*, 1987; Walker *et al.*, 2010).

On the other hand, some cross-modal mappings are not so stable. For example, Smith and Sera (1992) found that the mapping between size and darkness changes over time. Very young children consistently map light grey to small objects and dark grey to big objects, but this mapping gradually weakens and adults no longer show a consistent preference. This indicates that some

Address for correspondence: Shakila Shayan, Max Planck Institute for Psycholinguistics, P.O. Box 310, 6500 AH Nijmegen, The Netherlands; e-mail: shakila.shayan@mpi.nl

*Melissa Bowerman died in October 2011.

cross-modal mappings may be open to learning. In this paper we focus on how children and adults in three different language communities match sound to spatial extent, specifically pitch to thickness. We focus on this particular cross-modal association because there is evidence for both the hypotheses that cross-modal associations in this domain are a matter of mere convention and that they are grounded in universal principles.

Cross-linguistic comparison shows that language communities differ in the metaphors they use to describe pitch. In English and German, for example, sound is mapped onto vertical space, so we speak of 'high pitch' versus 'low pitch'. In contrast, in Farsi (spoken in Iran) and Turkish (spoken in Turkey) pitch is described with a 'thickness' metaphor: 'thin' for high-pitched sounds and 'thick' for low-pitched sounds (Shayan, Ozturk & Sicoli, 2011; Levinson & Majid, 2007). The existence of different metaphors suggests that pitch can be mapped to space in alternative ways, and perhaps none of these is privileged for human cognizers (Dolscheid, Shayan, Majid & Casasanto, 2013; cf. Eitan & Timmers, 2010).

Variable mappings may not be purely cultural constructs, but could also pick up on variable experiences with the world and our own bodies. The mapping of pitch onto vertical space (low and high) could be related to the area of the body that resonates with pitch range, in particular to the position of the larynx, which produces higher pitch when raised and lower pitch when lowered. In contrast, the thick–thin metaphor could be grounded in associations involving real-world objects, e.g. big people and animals typically produce lower-pitched sounds than children and small animals, and thicker strings on instruments produce lower-pitched tones than thinner strings. Given these real-world associations between pitch and both height and thickness, perhaps all people show similar cross-modal associations, either as the outcome of learning or due to innate biases (Shepard, 1994; Maurer & Mondloch, 2006).

We compared child and adult speakers of Farsi and Turkish (languages with the 'thick–thin' pitch metaphor) and German (a language without this metaphor) in a cross-modal matching task. Previous research shows that there is considerable development in cross-modal associations. Two-year-olds fail to make some auditory–visual correspondences that 3- and 4-year-olds make easily (Smith & Sera, 1992). So in this study we focused on 2- to 5-year-old children in order to explore the developmental trajectory of the pitch-to-thickness mapping. We focus on pitch and thickness since to our knowledge there are no previous studies examining this mapping in children (in contrast to the association of

pitch with height or size; cf. Marks *et al.*, 1987; Walker *et al.*, 2010; Wagner, Winner, Cicchetti & Gardner, 1981).¹

If the mapping of pitch to thickness is hardwired, and so available to all, then speakers of our three languages should behave in the same way regardless of age. That is, everyone should match thick objects to low-pitched sounds and thin objects to high-pitched sounds. If, in contrast, it is language that promotes this mapping, then Turkish and Farsi children and adults should map pitch to thickness consistently but German children and adults should do so only at chance levels. Finally, if the factor that is crucial to this mapping is experience with regular nonlinguistic associations between pitch and thickness, then adult speakers of all three languages should map consistently across modalities, but young children might not.

The study

Our study consisted of three tasks. A Word Comprehension study designed to determine whether participants understood the spatial and pitch-related meanings of the relevant words in their language, and two cross-modal association tasks: a Linguistic Similarity Matching task and a Nonlinguistic Similarity Matching task.

The purpose of the Linguistic Similarity Matching task was to establish whether Farsi and Turkish speakers map in the way predicted by their language, and also to discover whether German speakers could map in the same way even though their language does not promote this mapping. Participants heard a description of one of two pitches (high/low) or of one of two objects differing in thickness (snakes), and then matched this referent to one of the two entities on the other dimension. We expected that Farsi and Turkish speakers would successfully map the high tone to the thin snake and the low tone to the thick snake, and vice versa, since the same words are used for both dimensions. Note, however, that participants could not solve the task simply by matching the word used for one dimension to the same word on the other dimension; they had to make a cross-modal association between the described sound (or snake) to an actual snake (or sound). This task might be difficult for the German speakers, since German uses different words for the two modalities.

¹ 'Thick' stimuli are also bigger than 'thin' stimuli, so this manipulation could also be taken to be one of size but, critically, the size variation is only in one dimension.

The purpose of the Nonlinguistic Similarity Matching task was to determine whether participants could perform the cross-modal mapping even when the linguistic cues (for the Turkish and Farsi speakers) were eliminated. Participants were presented with a target stimulus on one dimension (for example, a high- or low-pitched sound) and asked to indicate which of the two stimuli on the other dimension (for example, a thin or thick snake) matched it. Since no linguistic labels were used, there was no explicit cue to advantage the Farsi and Turkish speakers over the German speakers.

The two matching tasks were run as a within-subject study. They were conducted in two sessions separated by a week, with the order of the tasks counter-balanced across participants. Prior to both Linguistic and Nonlinguistic tasks a familiarization task was performed which involved matching animals with their sounds. The purpose of the familiarization task was to ensure that children understood the cross-modal mapping task. The Word Comprehension task followed the Linguistic Similarity Matching task and was performed in the same session.

Method

Participants

Child and adult speakers of Turkish, Farsi and German participated in the experiment. The Turkish-speaking participants were 11 adults and 31 children (age range 2;4–4;11), all living in Turkey. The Farsi-speaking participants were 12 adults and 38 children (age range 2;7–5;0) living in Iran. All Farsi-speaking children participated in the Nonlinguistic Similarity Matching task, but only 17 finished the Linguistic Similarity Matching task due to a chickenpox epidemic. It was not feasible to make a repeat trip to Iran to collect the missing data. The German-speaking participants were 12 adults and 44 children (age range 2;4–4;11) living in Germany and the Netherlands. Children were recruited from daycares in Istanbul, Tehran, and Kleve, and tested on the premises in a separate room. Except for the 12 German-speaking adults who all lived in the Netherlands and had some knowledge of Dutch,² all the speakers were monolingual.

Material

For the thickness dimension, two snakes were cut out of cardboard and patterned with snake-skin. The snakes

were both 1 meter long; one was 2 cm wide and the other 12 cm wide. For the pitch dimension, there were two monotonic tones, one high-pitched (512 Hz) and one low-pitched (256 Hz). The frequencies were selected to eliminate variations in tone chroma while varying pitch height. These were created using the computer program Audacity. Each sound stimulus consisted of two 700 ms repetitions of the tone, separated by 250 ms of silence. High and low sounds were placed on two separate digital sound recorders connected to separate speakers, so they could be played from different locations. The speakers were placed in two cube-shaped houses made of cardboard. In addition, a set of six cut-out animals (cow, duck, monkey, cat, rabbit, dog) and three animal sound files (duck, cat, dog) were used in the familiarization task.

Procedure

Familiarization task

Before beginning the experimental trials, all children participated in a familiarization phase. The familiarization trials had the same structure as experimental trials. In half of the trials, children had to match a picture to a sound. Participants saw two cut-out animals and heard an animal call (e.g. bark). The child was told the animal in the house was calling for its friend who was very much like him, and wanted to be his friend. The child was asked which of the two animals was the friend. Once the child could answer at least two trials correctly, she/he was told they would play a new game.

The children then matched a sound to a picture. One of the animals was placed on the table and the child was told that this animal was looking for its friend and could only find it by listening to the animals in the houses. The two animal calls were played and the child had to point to the friend's house. Trials were repeated until the child could correctly perform the task. If a child failed the familiarization trials, they did not continue to the experimental trials. Two Farsi-speaking and three German-speaking children were excluded in this phase.

Linguistic Similarity Matching task

In this task participants were presented with an adjective describing the value an absent referent had on one dimension (e.g. thickness of an unseen snake) and were asked to match this referent to one of two perceptually present choices on the other dimension (e.g. high or low sound). There were two conditions. In the Thickness-to-Pitch trials participants were told there was a 'thick' ('thin') snake in the room looking for its friend, and that the friend was in one of the two houses and was very

²Pitch is described in Dutch, as in German, with the height metaphor.

much like this snake (see Table 1). Participants were instructed to listen to sounds coming from the houses to find out which house belonged to the friend. The experimenter then played the tones one after another from the two houses, and the participant had to point to a house. The left-right location of the tones was counter-balanced across trials.

In the Pitch-to-Thickness condition participants were told that there was a snake in one of the houses whose voice was ‘high/low’ (German) or ‘thin/thick’ (Farsi and Turkish) and it was calling for its friend, who was very much like it. Next the two cardboard snakes were placed on the table in front of the participant, one in front of the other. The participant was told that one of these snakes was the friend being sought, and asked to point to it. The left-right position of the sounds was counter-balanced across participants, as was the front-back position of the snakes. There were two trials in each condition (Figure 1), each repeated four times. Four random orders of the trials

Table 1 Adjectives describing high pitch, low pitch, thick and thin in Farsi, Turkish, and German

	High pitch	Low pitch	Thin	Thick
Farsi	nāzok	koloft	nāzok	koloft
Turkish	ince	kalmn	ince	kalmn
German	hoch	tief	dünn	dick

(a)

Exemplar	Choices	
Thick	256 Hz	512 Hz
Thin	256 Hz	512 Hz
Low		
High		

(b)

Exemplar	Choices	
	256 Hz	512 Hz
	256 Hz	512 Hz
256 Hz		
512 Hz		

Figure 1 Illustration of the four trial types in the (a) Linguistic Similarity Matching task and (b) Nonlinguistic Similarity Matching task.

were administered. Before beginning the experimental trials participants were told that they were going to play a game involving snakes looking for their friends, and they saw both snakes and heard both tones.

Nonlinguistic Similarity Matching task

The procedure and setup were similar to the Linguistic Similarity Matching task, except that stimuli were never described in words: the sounds were simply played and the cut-out snakes were shown. In the Thickness-to-Pitch condition, one snake was placed equidistant from the two houses (Figure 2a). Participants first heard that the snake was looking for its (very similar) friend, then listened to two tones coming from the two houses, and finally had to point to the friend’s house. In the Pitch-to-Thickness condition a sound was played and the choice was between the two snakes (Figure 2b). Participants were told that there was a snake in one of the houses calling for its friend. The experimenter then played a sound at one of these locations. Next both snakes were placed on the table in front of the participant, who was then asked to point to the snake that was the friend.

Word Comprehension task

Participants were separately assessed on their comprehension of the target adjectives in their language, for both their spatial and pitch senses. For the spatial stimuli, participants were presented with the two snakes and were asked: ‘which one is the “thick”/“thin” one’. For the sound stimuli, participants heard the two tones and were asked: ‘which one is the “low-pitched”/“high-pitched” one’ (‘thick’/‘thin’ for the Farsi and Turkish speakers, ‘low’/‘high’ for the German speakers; Table 1). Each adjective was tested twice in each modality.

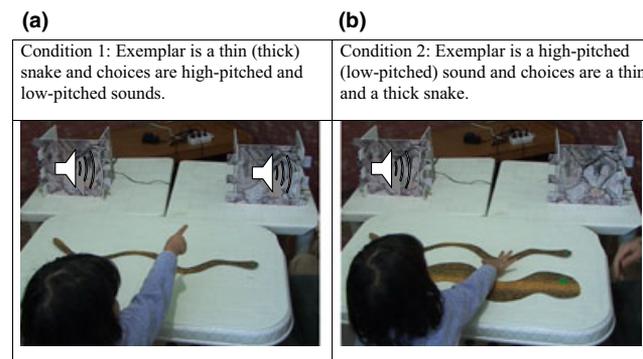


Figure 2 Two conditions in the Nonlinguistic Similarity Matching task.

Results

Word Comprehension task

First we established whether participants understood the spatial and sound senses of the words we tested (Table 2). Adults performed at ceiling in all three languages. For each adjective in each language, a one-sample t-test compared the mean of children's responses against chance. Farsi and Turkish children performed significantly better than chance on all words. German children were significantly better than chance on the spatial words, but at chance on the pitch words. (Remember, German uses a different spatial metaphor for sound from Farsi and Turkish.)

Linguistic Similarity Matching task

Responses were considered 'correct' if they followed the Farsi/Turkish pitch-thickness mapping associating a high sound with a thin snake and a low sound with a thick snake. We first analyzed the data by binning children into two groups (older and younger) using median-split on age in months. There was no significant change in behavior between younger and older children, nor was there a correlation between cross-modal mappings and age in months. So we pooled the child data. Adult speakers of all three languages as well as Farsi and Turkish children mapped correctly, but German children were at chance (Figure 3a). A 3(Language) \times 2(Age: children vs. adults) \times 2(Condition: Pitch-to-Thickness, Thickness-to-Pitch) mixed ANOVA showed no main effect of Condition, $F(1, 118) = 2.09, p > .1, \eta_p^2 = .01$, nor any interaction between Language and Condition, $F(2, 118) = .02, p > .9, \eta_p^2 < .001$, or Age and Condition, $F(1, 118) = 2.4, p > .1, \eta_p^2 = .02$. So the two conditions were collapsed and the data were submitted to a 3(Language) \times 2(Age) ANOVA. There was

a main effect of both Age, $F(1, 118) = 125.5, p < .001, \eta_p^2 = .51$, Language, $F(2, 118) = 17.6, p < .001, \eta_p^2 = .23$, and an Age \times Language interaction, $F(2, 118) = 11.83, p < .001, \eta_p^2 = .19$. Independent pairwise comparisons between child groups revealed a difference between German and Farsi and between German and Turkish (both $p < .001$), but not between Farsi and Turkish ($p > .1$). One-sample t-tests for the children in each language group revealed that Farsi and Turkish children were significantly better than chance, $t(16) = 8.83, p < .001$, and $t(30) = 8.30, p < .001$, respectively, but German children were not, $t(40) = .14, p > .8$.

Adult speakers of all three languages were able to perform the linguistic cross-modal mapping between pitch and thickness, and so were the Farsi and Turkish children, but not the German children. This pattern of results suggests that mapping is promoted both by language and by nonlinguistic experience. German does not have a thick–thin metaphor for sound, so the ability of German-speaking adults to perform the mapping cannot be due to language. But the ability is also not likely to be inborn, because German children were only at chance on this task. So presumably the German adults had enough nonlinguistic experience with typical correlations between size and pitch to make this association, but the German children had not.

Nonlinguistic Similarity Matching task

Just as for the Linguistic Similarity Matching task, responses were considered correct when participants mapped between a high sound and thin snake and between a low sound and thick snake. And, as before, preliminary analyses showed no developmental change within the child groups. We therefore compared children directly with adults. A comparison of Figure 3b with Figure 3a shows that the results of the Nonlinguistic Similarity Matching task are very similar to those of the

Table 2 Percentage of correct responses in the Word Comprehension task for each stimulus item in each language. Values marked with an asterisk were significantly better than chance ($p < .01$)

Stimuli	High pitched tone		Low pitched tone		Thin snake		Thick snake	
<i>a) Farsi</i>								
Adjectives	nāzok		koloft		nāzok		koloft	
	% correct	<i>t</i> -value	% correct	<i>t</i> -value	% correct	<i>t</i> -value	% correct	<i>t</i> -value
	67	$t(30) = 2.8^*$	82	$t(30) = 6.9^*$	72	$t(33) = 3.2^*$	75	$t(30) = 4.1^*$
<i>b) Turkish</i>								
Adjectives	ince		kalin		ince		kalin	
	% correct	<i>t</i> -value	% correct	<i>t</i> -value	% correct	<i>t</i> -value	% correct	<i>t</i> -value
	80	$t(30) = 6.1^*$	74	$t(30) = 4.3^*$	66	$t(30) = 2.7^*$	84	$t(30) = 7.9^*$
<i>c) German</i>								
Adjectives	hoch		tief		dünn		dick	
	% correct	<i>t</i> -value	% correct	<i>t</i> -value	% correct	<i>t</i> -value	% correct	<i>t</i> -value
	47	$t(39) = -.49$	56	$t(39) = 1.1$	70	$t(39) = 6.4^*$	80	$t(39) = 6.4^*$

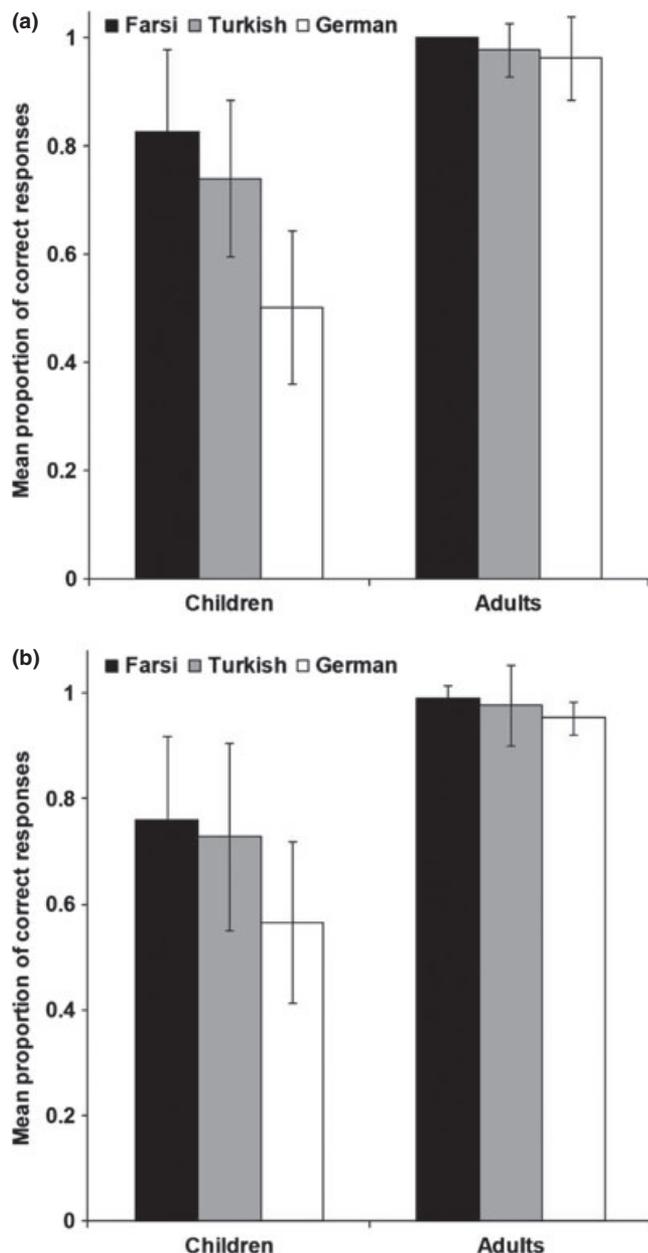


Figure 3 Mean proportion of correct responses (bars show standard deviation) for the (a) Linguistic Similarity Matching task and (b) Nonlinguistic Similarity Matching task, where 'correct' is mapping thick to low pitch and thin to high pitch.

Linguistic Similarity Matching task, with all the adults as well as the Farsi- and Turkish-speaking children performing well on both tasks and the German-speaking children at chance.

A 3(Language) \times 2(Age) \times 2(Condition) mixed ANOVA showed no effect of Condition, $F(1, 139) = .86$, $p > .3$, $\eta_p^2 = .006$, and no interaction between Condition

and Language, $F(2, 139) = .71$, $p > .48$, $\eta_p^2 = .01$, or Condition and Age, $F(1, 139) = .001$, $p > .9$, $\eta_p^2 < .001$. So conditions were collapsed and the data were submitted to a 3(Language) \times 2(Age) ANOVA. There was a main effect of Age, $F(1, 139) = 103.56$, $p < .001$, $\eta_p^2 = .43$, Language, $F(2, 139) = 6.46$, $p < .001$, $\eta_p^2 = .08$, and an Age \times Language interaction, $F(2, 139) = 3.15$, $p < .04$, $\eta_p^2 = .04$. Pairwise comparisons revealed a difference between German and Farsi ($p < .001$) and German and Turkish ($p < .001$), but not between Farsi and Turkish ($p > 1$). One-sample t-tests showed that performance was significantly better than chance for Farsi and Turkish children, $t(38) = 10.16$, $p < .001$ and $t(30) = 9.24$, $p < .001$, respectively, but not for German children, $t(40) = 1.90$, $p = .06$.

Our linguistic and nonlinguistic tasks were conducted in two separate sessions a week apart. Half the participants carried out the linguistic task in the first session and half in the second session. To control for the possibility that children who performed the linguistic task first had carried over some information to the nonlinguistic task, we looked separately only at the subgroup of children who had performed the nonlinguistic task first. A one-way ANOVA on their responses in the Nonlinguistic Similarity Matching task revealed a main effect of language, $F(2, 72) = 16.86$, $p < .001$. Post-hoc analyses showed no difference between the Farsi and Turkish children ($p > .6$), but a difference between the German children and both other groups ($p < .001$). This outcome – based only on the children who performed the nonlinguistic task first – indicates that Turkish and Farsi children are able to map cross-modally across pitch and thickness even when they receive no explicit linguistic cuing.

Relationship between Linguistic and Nonlinguistic tasks

The previous analyses show that Turkish and Farsi children map between pitch and thickness in both the linguistic and nonlinguistic tasks, whereas the German children do not, consistent with the hypothesis that language promotes this cross-modal mapping. If so, then there should also be a positive correlation between the linguistic and nonlinguistic tasks for Turkish and Farsi children, and there is, $r(45) = .66$, $p < .0001$.

Discussion

We have shown that cross-modal associations between pitch and thickness are equally available to adult speakers of diverse languages, but not necessarily to

children. In both the Linguistic Similarity Matching task and the Nonlinguistic Similarity Matching task, adult speakers of all three language groups – Farsi, Turkish, and German – could map between pitch and thickness according to the Farsi/Turkish metaphor (an association between high pitch and thin, and between low pitch and thick). But children's performance was language specific: Turkish- and Farsi-speaking children consistently mapped between high pitch and thin and low pitch and thick, while German-speaking children's mappings were at chance.

In the real world, pitch often correlates with spatial extent, with thinner or smaller animals and objects producing higher-pitched sounds than thicker or larger ones. Although this correlation is apparently available to German-speaking adults, it is not yet available to German children. This result resonates with previous studies by Marks *et al.* (1987), who showed that children could not systematically map size (big vs. small) to pitch until they were 13 years old.

These results suggest that the association between size/thickness and pitch is learned. This is consistent with other findings that indicate that cross-modal mappings between space and pitch vary as a function of culture. When asked to sing back a note while viewing an irrelevant spatial stimulus, Farsi speakers sing the same note higher when viewing a thinner (vs. thicker) line but their performance is not affected by a line placed higher (vs. lower) in space. Dutch speakers, however, show the opposite pattern. They sing back the same note higher when there is an irrelevant high spatial stimulus, but are uninfluenced by thick stimuli (Dolscheid *et al.*, 2013). In this study we found that German speakers, unlike Dutch speakers in the Dolscheid *et al.* study, could make a thickness-to-pitch mapping – presumably because this task was very simple.

Our study suggests that the thickness-to-pitch mapping can be learned, and it can be learned on a purely nonlinguistic basis (cf. the German adults who have mastered the mapping even though their language does not promote it). But the real-world co-occurrences between pitch and size/thickness are apparently not salient or frequent enough to be learned in the first few years of life purely on the basis of nonlinguistic experience (cf. the German children).

Farsi and Turkish children receive an additional boost from their languages: the cross-modal association between pitch and thickness that is conventional in these languages apparently enhances the salience of the mapping so that even 3-years-olds, who are usually poor in relational structure mapping (Gentner, 2003), succeed on our nonlinguistic task. Moreover, there is a strong positive correlation between the Linguistic and

Nonlinguistic tasks for the Turkish and Farsi children consistent with hypothesis that language helps promote this cross-modal mapping. Of course, correlation does not imply causation. There could be some third environmental or cultural factor that underlies the pitch-to-thickness mapping for the Turkish and Farsi children in this study. Perhaps Turkish and Farsi children have more experience with snakes and sounds than German children. Based on ethnographic data, this seems highly implausible, but it cannot be ruled out without further systematic investigation, including a more comprehensive sample of spatial stimuli manipulating thickness. On the other hand, we do know that language can play a causal role in establishing a mapping between thickness and pitch. After training adult Dutch speakers to use Turkish- and Farsi-like 'thickness' metaphors, participants were able to map pitch-to-thickness in a task where they were previously unable to make the mapping (Dolscheid *et al.*, 2013). Taken together these findings suggest that language can encourage cross-modal mappings.

Finally, we have argued that the pitch-to-thickness mapping is learned but can we rule out entirely the possibility that the cross-modal mapping is innate? From this data alone we cannot. The German children do not make the pitch-to-thickness mapping, arguing against innate representation of this cross-modal correspondence, but it is possible that they did have knowledge of this mapping at birth but have since lost it (cf. Hespos & Spelke, 2004). Only studies with very young infants can address the question of innate representations of cross-modal correspondences.

To summarize, the current study suggests that the pitch-to-thickness mapping is learned, and it can be learned faster if the language you speak promotes that mapping.

Acknowledgements

This research was funded by the Max Planck Gesellschaft. We would like to thank members of the Categories across Language and Cognition project at the Max Planck Institute for Psycholinguistics for critical feedback and discussion, especially Disa Sauter. Thanks also to Marcel Braun, Lisa Dickmann, Amelie Theiler, Katja Towet, and Luzia Unverricht, who ran the German participants. Special thanks to daycare teachers and coordinators in Kleve, Istanbul, and Tehran for their enormous help, and in particular to Zahra Farmani, Mahvash Fouladi, Fatemeh Soltani, and Safoura Tofangchi for their help in finding daycare centers in Tehran.

References

- Dolscheid, S., Shayan, S., Majid, A., & Casasanto, D. (2013). The thickness of musical pitch: Psychophysical evidence for linguistic relativity. *Psychological Science*, **24**, 613–621.
- 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**, 405–422.
- Gentner, D. (2003). Why we're so smart. In D. Gentner & S. Goldin-Meadow (Eds.), *Language in mind: Advances in the study of language and thought* (pp. 195–235). Cambridge, MA: MIT Press.
- Hespos, S.J., & Spelke, E.S. (2004). Conceptual precursors to language. *Nature*, **430**, 453–456.
- Levinson, S.C., & Majid, A. (2007). The language of sound. In A. Majid (Ed.), *Field manual, Volume 10* (pp. 29–31). Nijmegen: Max Planck Institute for Psycholinguistics.
- Marks, L.E. (1974). On associations of light and sound: The mediation of brightness, pitch, and loudness. *American Journal of Psychology*, **87**, 173–188.
- Marks, L.E. (1978). *The unity of the senses: Interrelations among the modalities*. New York: Academic Press.
- 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., Hammeal, R.J., & Bornstein, L.B. (1987). Perceiving similarity and comprehending metaphor. *Monographs of the Society for Research in Child Development*, **52** (1, Serial No. 215).
- Maurer, D., & Mondloch, C.J. (2006). The infant as synesthete? *Attention and Performance*, **21**, 449–471.
- Mondloch, C.J., & Maurer, D. (2004). Do small white balls squeak? Pitch–object correspondences in young children. *Cognitive, Affective, & Behavioral Neuroscience*, **4**, 133–136.
- Shayan, S., Ozturk, O., & Sicoli, M.A. (2011). The thickness of pitch: Crossmodal metaphors in Farsi, Turkish and Zapotec. *The Senses & Society*, **6**, 96–105.
- Shepard, R.N. (1994). Perceptual-cognitive universals as reflections of the world. *Psychonomic Bulletin & Review*, **1**, 2–28.
- Smith, L.B., & Sera, M.D. (1992). A developmental analysis of the polar structure of dimensions. *Cognitive Psychology*, **24**, 99–142.
- 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., Maststock, K., Slater, A., & Johnson, S. (2010). Preverbal infants' sensitivity to synaesthetic cross-modality correspondences. *Psychological Science*, **21**, 21–25.