Check for updates

Received: 15 May 2022 | Revised: 8 September 2022 | Accepted: 18 October 2022

DOI: 10.1111/desc.13341

Developmental Science WILEY

RESEARCH ARTICLE

Children's associations between space and pitch are differentially shaped by language

Sarah Dolscheid[1] | Simge Çelik[2] | Hasan Erkan[3] | Aylin Küntay[4] | Asifa Majid[5]

[1]University of Cologne, Department of Rehabilitation and Special Education, Cologne, Germany

[2]Heidelberg University, Heidelberg, Germany

[3]Radboud University, Nijmegen, The Netherlands

[4]Koç University, Istanbul, Turkey

[5]University of Oxford, England, UK

Correspondence
Asifa Majid, Department of Experimental Psychology, University of Oxford, United Kingdom.
Email: asifa.majid@psy.ox.ac.uk

Abstract

Musical properties, such as auditory pitch, are not expressed in the same way across cultures. In some languages, pitch is expressed in terms of spatial height (high vs. low), whereas others rely on thickness vocabulary (thick = low frequency vs. thin = high frequency). We investigated how children represent pitch in the face of this variable linguistic input by examining the developmental trajectory of linguistic and non-linguistic space-pitch associations in children who acquire Dutch (a height-pitch language) or Turkish (a thickness-pitch language). Five-year-olds, 7-year-olds, 9-year-olds, and 11-year-olds were tested for their understanding of pitch terminology and their associations of spatial dimensions with auditory pitch when no language was used. Across tasks, thickness-pitch associations were more robust than height-pitch associations. This was true for Turkish children, and also Dutch children not exposed to thickness-pitch vocabulary. Height-pitch associations, on the other hand, were not reliable—not even in Dutch-speaking children until age 11—the age when they demonstrated full comprehension of height-pitch terminology. Moreover, Turkish-speaking children reversed height-pitch associations. Taken together, these findings suggest thickness-pitch associations are acquired in similar ways by children from different cultures, but the acquisition of height-pitch associations is more susceptible to linguistic input. Overall, then, despite cross-cultural stability in some components, there is variation in how children come to represent musical pitch, one of the building blocks of music.

KEYWORDS
cross-cultural, development, language, linguistic relativity, music terminology, pitch

Research Highlights

- Children from diverse cultures differ in their understanding of music vocabulary and in their nonlinguistic associations between spatial dimensions and auditory pitch.

This is an open access article under the terms of the Creative Commons Attribution-NonCommercial License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited and is not used for commercial purposes.
© 2021 The Authors. *Developmental Science* published by John Wiley & Sons Ltd.

14677687, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/desc.13341 by Koc University, Wiley Online Library on [22/02/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

- Height-pitch mappings are acquired late and require additional scaffolding from language, whereas thickness-pitch mappings are acquired early and are less susceptible to language input.
- Space-pitch mappings are not static from birth to adulthood, but change over development, suggesting music cognition is shaped by cross-cultural experience.

1 | INTRODUCTION

Auditory pitch plays a central role in music. In particular, successive changes in pitch direction (i.e., contour) form the essence of a melody. To describe changes in pitch, speakers of English inevitably use spatial terminology (Pratt, 1930). For instance, pitch can be described as 'high' or 'low', a melody can 'rise' or 'fall', and a musical scale can 'ascend' or 'descend'. English speakers thus describe auditory pitch through spatial terminology, especially spatial height. However, languages vary considerably in how they express musical pitch (e.g., Eitan & Timmers, 2009; Majid et al., 2018; Shayan et al., 2011). For example, pitch polarity is described as light versus heavy (Kpelle, Liberia) or young versus old (Suyá, Brazil) in some languages and cultures (Eitan & Timmers, 2009). Even when using spatial terminology, there are differences across languages. The Austronesian language 'Are'are, for instance, expresses pitch in terms of spatial height, but a high-pitched tone is 'low' or 'down' (Zemp & Malkus, 1979). Languages like Farsi (Iran), Turkish (Turkey), and Zapotec (Mexico), on the other hand, do not refer to height but spatial thickness, with high pitches described as 'thin' and low pitches as 'thick' (Shayan et al., 2011). As these examples illustrate, languages across the world differ substantively in their pitch vocabularies.

There is evidence that differences in pitch terminology can also influence the way people represent pitch, even in entirely nonlinguistic tasks (Dolscheid et al., 2013, 2020; Fernandez-Prieto et al., 2017). In one study, Dutch speakers (who talk about pitch in terms of spatial height) and Farsi speakers (who talk about pitch in terms of thickness) were asked to sing back tones they heard in the presence of irrelevant visuo-spatial information, that is, lines that varied in either height or thickness. Whereas Dutch speakers' singing was influenced by spatial height but not thickness, the reverse was true for Farsi speakers, thus demonstrating non-linguistic pitch representations differ in language-specific ways (Dolscheid et al., 2013). Converging evidence comes from another study that showed English speakers (who preferentially use height-pitch terminology) display stronger non-linguistic associations between spatial height and pitch than speakers of Catalan or Spanish (who make less consistent use of height-pitch terminology) (Fernandez-Prieto et al., 2017). So differences in music terminology—specifically pitch vocabularies—can affect nonlinguistic associations between space and auditory pitch.

Critically, although language shapes space-pitch associations, it does not establish them in the first place. Even young infants possess implicit associations between spatial dimensions and pitch (Dolscheid et al., 2014; Jeschonek et al., 2013; Pietraszewski et al., 2017; Walker et al., 2010, 2018; but see Lewkowicz & Minar, 2014). For example, 4-month-old infants prefer to look at a visual display when a ball moving upward is accompanied by a rising rather than falling pitch, and vice versa for a ball moving downwards (Dolscheid et al., 2014; Walker et al., 2010). Infants also look longer at displays when a rising pitch is paired with a tube decreasing in thickness rather than increasing, and vice versa for falling pitch (Dolscheid et al., 2014). Thus, 4-month-old infants are sensitive to both height-pitch and thickness-pitch mappings, suggesting these associations are present prelinguistically.

The existence of early space-pitch mappings in infants seems to contrast with the language-specific associations evident in adults, begging the question of what happens between infancy and adulthood. Do children who acquire different pitch terminologies also differ in their nonlinguistic space-pitch associations? If so, when and how are children's pitch representations shaped by language? These questions feed directly into larger debates about cultural influences on music perception. On the one hand, music seems to be universal (Mehr et al., 2019; Trehub, 2000), and even infants with minimal exposure to music show striking similarities to adult listeners, suggesting the presence of universal perceptual biases in relation to music (for an overview see, e.g., Trehub, 2000). However, there is also substantial cross-cultural variation (Nettl, 2000; Trehub et al., 2015), as well as diversity in how children come to perceive music (Hannon et al., 2012; Trehub et al., 2008; Weiss et al., 2020). So a key question is to what extent fundamental properties of music perception are shared across cultures and over development and to what extent they are susceptible to variations of experience. In the present study, we seek to address this by focusing on a particular type of experience—namely, differential exposure to music vocabularies across languages. Specifically, we investigate whether cross-linguistic differences in pitch terminologies modulate how children represent auditory pitch.

1.1 | The development of associations between auditory pitch and space

In a previous cross-linguistic study, Shayan et al. (2014) examined children's associations between thickness and pitch by comparing cohorts of Turkish, Farsi, and German-speaking children. Two- to 5-year-olds saw a cardboard snake who was either thin or thick (i.e., physically narrower or wider) and were told the snake was looking for its friend in one of two houses. Children then listened to two tones—a high and low pitch sound—emanating successively from each house, and had to decide where the snake's friend lived by pointing to the respective house. Crucially, stimuli were never described in words.

14677687, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/desc.13341 by Koc University, Wiley Online Library on [22/02/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

Nonetheless, children differed in their mapping choices: Farsi and Turkish children associated spatial thickness and pitch consistently, but the same-aged German children—who are not exposed to thickness-pitch terminology—did not (Shayan et al., 2014). This suggests linguistic terminology promotes the development of nonlinguistic space-pitch associations.

Other studies appear at odds with Shayan and colleagues' conclusions, however. Starr & Srinivasan (2018), for example, found English speaking children not only associated pitch with spatial height, but also displayed consistent associations between pitch and thickness (see also Starr et al., 2021). This was true even though English children are not exposed to thickness-pitch vocabulary in their input language, challenging the idea that linguistic terminology is critically involved in children's pitch representations.

Given these inconsistent findings, the role of language in shaping children's space-pitch associations remains unclear. In the present study, we build on previous work by addressing some limitations of earlier studies. We move beyond the work of Shayan et al. (2014) by examining not only thickness-pitch associations, but also height-pitch associations, as Starr & Srinivasan (2018) did in their study of English-speaking children. In addition, we test children from two different communities—Dutch and Turkish—who are exposed to different pitch vocabularies (i.e., height vs. thickness), thus going beyond the study of a single WEIRD community (Henrich et al., 2010; Kidd & Garcia, 2022).

If language influences nonlinguistic associations between space and pitch, Dutch and Turkish children's associations should differ in accordance with the pitch vocabularies of each language. Alternatively, if language does not (yet) influence associations between space and pitch, Dutch and Turkish children's performance should be similar for the two types of space-pitch associations. Since children seem to acquire height-pitch terminology remarkably late (e.g., Andrews & Madeira, 1977; Costa-Giomi & Descombes, 1996; Durkin & Shire, 1990; Durkin & Townsend, 1997; Webster & Schlentrich, 1982), we included a broad age range, testing children at 5, 7, 9, and 11 years to obtain more detailed insight into the developmental trajectories of space-pitch representations.

1.2 | The acquisition of vocabulary for auditory pitch

Another central goal of the study was to examine children's acquisition of pitch terminology itself. As has long been observed, children face difficulties in the acquisition of height-pitch terminology (Andrews & Madeira, 1977; Costa-Giomi & Descombes, 1996; Durkin & Townsend, 1997; Webster & Schlentrich, 1982; but see Stalinski et al., 2008; Starr & Srinivasan, 2018). For instance, Andrews & Madeira (1977) tested 6-, 7-, and 8-year-old children's ability to label tones as high or low and found only the oldest cohort of 8-year-olds performed above chance (see Table 1, p. 285). In another task, the same children were asked to compare two tones and label the second as either higher or lower than the first. Again, only the oldest children were able to do this, suggesting the production of pitch terminology is difficult for the child learner.

Similar difficulties have been observed for children's comprehension of pitch terminology (Dolscheid et al., 2015; Sergeant, 1984; Shayan et al., 2014; although see Starr et al., 2021). For instance, when 4- to 9-year-olds were asked to find the button that played the highest/lowest sounds, all children performed at chance (Sergeant, 1984). Similarly, younger children (2- to 5-year-olds) did not reliably understand height-pitch terminology when asked which of two sounds was high or low i pitch (Shayan et al., 2014). So both communicating about pitch and understanding pitch terminology appears to be challenging for children.

A possible explanation for this difficulty is that pitch is relatively abstract: its perception is fleeting, limited to (largely) one channel (i.e., auditory, leaving aside felt vibrations for now), and not readily available for simultaneous comparison, making it harder for children to extrapolate relevant features for concept learning. Across languages, abstract vocabulary is learnt later than more concrete vocabulary (Braginsky et al., 2019), so the relatively abstract nature of auditory pitch may make it challenging for children to acquire. Consistent with this, a training study found 3- to 5-year-old English speaking children more easily acquired spatial than auditory meanings of novel words. However, the same study also found children were as adept mapping novel auditory meanings to space as they were mapping novel spatial meanings to sound, contrary to the predictions of a simple concreteness account (Starr et al., 2021).

Another possibility is that the polysemy of pitch terminology (i.e., the multiple meanings of the words *high* and *low*) causes problems for the child learner (Costa-Giomi & Descombes, 1996). Polysemous pitch vocabulary is said to evoke uncertainty because the child is faced with multiple meanings and so does not know how to interpret *high* and *low* in the context of musical pitch (e.g., Costa-Giomi & Descombes, 1996; Durkin & Townsend, 1997). In line with this, French speaking children are better able to describe musical tones when trained with antonyms exclusively dedicated to pitch (*aigu* and *grave*) compared to polysemous expressions (*haut* 'high' and *bas* 'low') used for both spatial height and auditory pitch (Costa-Giomi & Descombes, 1996).

However, other evidence suggests polysemy cannot be the only explanation for the difficulties children have with pitch vocabulary. A polysemy account would predict children are confused about the multiple potential mappings of height-pitch terminology and struggle to map domains consistently (e.g. Durkin & Townsend, 1997). In contrast, children make regular errors (e.g., Dolscheid et al., 2015; Sergeant, 1984): in one study around half the tested children reversed 'high' and 'low' in the context of musical pitch (Hitchcock, 1942, as cited in Sergeant, 1984). A later study replicated this reversal, finding most 5-year-olds associate 'high' with a low-pitch tone and 'low' with a high-pitch tone (Dolscheid et al., 2015). So children systematically reverse height-pitch polarity, suggesting they are not simply confused by the polysemy of pitch vocabulary, but instead they intuit a reversed alignment between spatial height and pitch. If true, then children's difficulties with height-pitch terminology might not be rooted in polysemy alone.

In the present study we sought to test this possibility by taking advantage of cross-linguistic differences between Dutch and Turkish.

14677687, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/desc.13341 by Koc University, Wiley Online Library on [22/02/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

While both Dutch and Turkish use polysemous space-pitch terminology, the two languages differ in the specific spatial dimension they use. Examining the acquisition of pitch vocabularies in both languages allows us to test which factors specifically contribute to children's difficulties with pitch terminology. If polysemy is the main problem, children should be equally late to acquire height-pitch and thickness-pitch terminology because both are polysemous. Alternatively, if children's difficulties are not due to polysemy alone, then height-pitch and thickness-pitch terminologies might be acquired in different ways.

1.3 | The present study

Accordingly, we investigated whether differences in experience (i.e., cross-linguistic variation in exposure to pitch terminologies) modulate how children represent auditory pitch—one important component of music. To do this, we took a cross-cultural and cross-sectional approach. In a battery of tasks, we tested children of different ages speaking Dutch and Turkish to examine: (a) how well they understood space-pitch terminology, and (b) how reliably they associated various spatial dimensions to pitch when no language was used. To yield a more comprehensive picture, we examined Dutch and Turkish children's comprehension of space-pitch terminology in their own language, and terminology not in their input language (i.e., comprehension of thickness-pitch terminology by Dutch children and height-pitch terminology by Turkish children). We also examined nonlinguistic height-pitch and thickness-pitch associations in both groups.

2 | METHODS

2.1 | Participants

Native Dutch (n = 80) and Turkish children (n = 80) participated in the experiments. Children were divided into four age groups with 40 children per group: 5-, 7-, 9- and 11-year-olds (see Table 1).

Dutch children were recruited from two primary schools in the Netherlands and Turkish children from several local primary schools in Turkey. All children had normal hearing and no known learning or language problems. According to a language background questionnaire administered to parents, the only language spoken at home was Dutch or Turkish, as dominant in each city where testing took place. Parents also filled out a musical background questionnaire, providing information about children's musical instrument training, number of years of training, and note reading abilities. Only five children out of 160 had some minimal musical training, and were therefore included in the study.

Informed consent was obtained from all participants, and ethical approval granted from the relevant universities and authorities prior to testing. Data and scripts for analyses are publicly available at OSF: https://osf.io/vgfzr/

2.2 | Procedure

The experiment was administered on a PC laptop using Presentation software (version 18.1, www.neurobs.com) and a Cedrus RB540 button box. Children were tested in silent (class)rooms in schools. Sounds were presented via headphones at approximately 60 dB-a. Participants sat in front of a laptop, next to the experimenter. Instructions were given verbally and on the laptop screen in the native language of the participant. All questions were read aloud by experimenters to mitigate against children's limited reading abilities. No space-pitch terminology was used during testing, nor was feedback supplied regarding the correctness of responses. The experiment lasted for approximately 35 min. For both linguistic and nonlinguistic tasks, the order of height and thickness conditions was counterbalanced across participants. To avoid spill-over effects from linguistic to nonlinguistic tasks, all participants were first tested in the nonlinguistic tasks. However, for exposition purposes, we report the findings for the linguistic tasks first.

2.3 | Materials

2.3.1 | Linguistic space-pitch comprehension tasks

To test participants' height and thickness pitch terminology, a linguistic space-pitch comprehension task was administered. Participants wore headphones and were presented with videos of two colored squares (square friends) that appeared simultaneously on the left and right side of the computer screen (see Figure 1). Each trial started with the left square 'singing' (accompanied by a wiggling move to indicate the square was making the sound) a single tone for two seconds followed by the right square 'singing' a single tone for two seconds. Three tone pairs (262 vs. 698 Hz, 262 vs. 523 Hz., 330 vs. 523 Hz) were presented four times. For each condition 12 trials were presented; half started with a high pitch and half with low pitch. In the height-pitch

TABLE 1 Demographic information by language and age group

	Turkish		Dutch	
Age group	*M* (*range*)	*n* (*female*)[1]	*M* (*range*)	*n* (*female*)
5-year-olds	5;2 (4;7–5;7)	20 (9)	5;7 (5;0–6;0)	20 (14)
7-year-olds	7;3 (7;0–7;7)	20 (12)	7;5 (6;10–7;7)	20 (10)
9-year-olds	9;2 (9;0–9;6)	20 (9)	9;6 (8;7–10;0)	20 (9)
11-year-olds	11;1 (10;8–11;5)	20 (10)	11;6 (10;11–12;1)	20 (13)

14677687, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/desc.13341 by Koc University, Wiley Online Library on [22/02/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

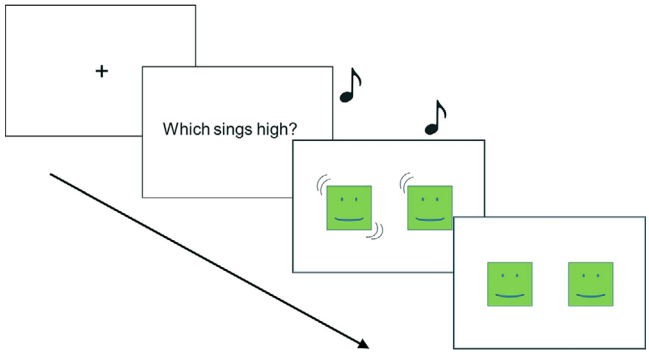


FIGURE 1 Example trial of the linguistic space-pitch comprehension task. Children were first asked a space-pitch question (e.g., *'Which sings high?'*) and then shown two 'square friends' sequentially producing a low-pitched versus high-pitched tone. They had to select which square produced the sound by pressing the corresponding button on a response box

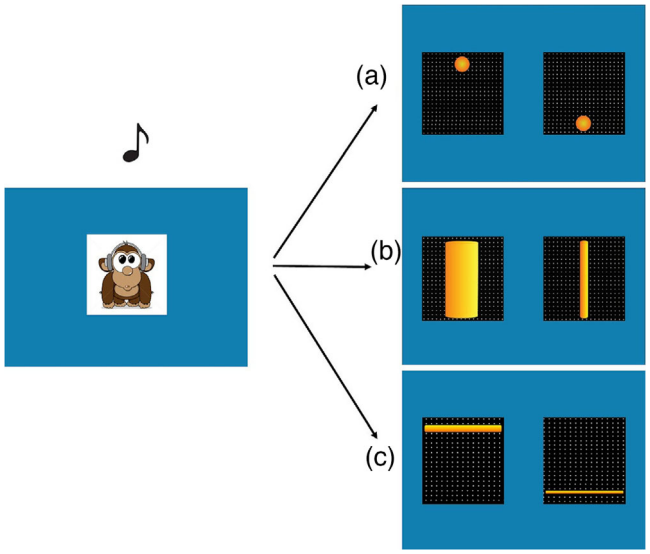


FIGURE 2 Example trial of the nonlinguistic space-pitch tasks. Children were first presented with one of two sounds (a high-pitch or low-pitch tone) before they had to indicate which image best matched the sound by pressing the corresponding button on a response box. Children were presented with: (a) balls that differed in height, (b) tubes that differed in thickness, or (c) juxtaposed spatial dimensions which created a conflict (i.e., a thick line presented high in space and a thin line presented low in space)

condition, children were asked *'Which sings high/low?'* before they listened to the tones. In the thickness-pitch condition, children were asked *'Which sings thick/thin?'*. The words tested in Dutch were *hoog/laag* and *dun/dik*, and in Turkish *yukarı/aşağı* and *ince/kalın* (high/low and thin/thick respectively). For each trial, children were asked to indicate which square belonged to the sound by pressing the corresponding button on a response box. Children were told they could listen to the sounds again in case they were not sure how to respond, in which case the whole trial was repeated. Responses on each trial were coded as 1 when they were in line with Dutch height-pitch and Turkish thickness-pitch terminology (i.e., *thin* = high pitch; *thick* = low pitch) and as 0 when they were not.

2.4 | Nonlinguistic space-pitch association tasks

Children first listened to one of two pure tones for 2 s (440 vs. 698 hertz). They were then presented with two spatial stimuli on the laptop screen. In the height-pitch condition, they were shown two orange balls (both approximately 2.2 cm diameter), presented side by side, with one ball at the top and the other on the bottom of a 12.5 × 11.3 cm grid of white dots on a black background (see Figure 2a). Children were asked to decide which of the two visual stimuli best matched the preceding sound by pressing the corresponding button (i.e., left vs. right) on a response box. The stimuli disappeared as soon as a button was pressed. If no response occurred within 30 seconds, the image disappeared and children were asked to select the corresponding image afterwards. Before the experimental trials began, there were two practice trials in order to familiarize participants with the task. Children were then presented with 8 experimental trials. No feedback was provided.

The thickness-pitch condition was identical to the height-pitch condition apart from the spatial stimuli. Instead of two balls, two vertical tubes of orange color and equal length (4.7 cm) were presented side by side, with one tube being relatively thin (1.2 cm width) and the other comparatively thick (5 cm width, see Figure 2b).

2.5 | Nonlinguistic space-pitch conflict task

The space-pitch conflict task was identical to the nonlinguistic tasks above except for the spatial stimuli. Children saw two horizontal lines side by side on the screen: one was relatively thick (9 mm width) and presented at the top of a 20 × 20 cm grid of white dots on a black background, and the other was relatively thin (1.5 mm width) and presented at the bottom of the grid (see Figure 2c). On each of 8 trials, children had to decide which best fit the tone played. Responses were coded as reflecting either a height-pitch or thickness-pitch mapping.

3 | RESULTS

3.1 | Linguistic tasks: Comprehension of space-pitch vocabulary

3.1.1 | Children's understanding of their native language pitch vocabulary

First, we examined how children mastered pitch terminology from their native languages (i.e., height-pitch for Dutch and thickness-pitch for Turkish). Overall, Dutch children's height-pitch comprehension was relatively poor, with on average 56% correct (collapsed across age groups). In order to test whether participants' performance was different from chance (50%), one-sample t-tests were conducted separately for each age group. Only 11-year-old Dutch children showed

14677687, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/desc.13341 by Koc University, Wiley Online Library on [22/02/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

above chance performance for height-pitch comprehension, 81% correct responses, $t(19) = 5.89$, $p < 0.001$, Cohen's $d = 1.32$. All other Dutch children performed at chance: 5-year-olds, 42%, $t(19) = -1.81$, $p = 0.09$, Cohen's $d = -0.40$; 7-year-olds, 48%, $t(19) = -0.23$, $= 0.82$, Cohen's $d = -0.05$; and 9-year-olds, 52%, $t(19) = 0.29$, $p = 0.78$, Cohen's $d = 0.06$.

In contrast, Turkish children had good comprehension of thickness-pitch, with on average 81% correct responses. Although the youngest age group, that is, 5-year-olds, performed at chance, 46% correct responses, $t(19) = 2.03$, $p = 0.06$, Cohen's $d = 0.45$; by 7 years, Turkish children were already performing better than chance on their native language musical vocabulary: 7-year-olds, 83%, $t(19) = 10.81$, $p < 0.001$, Cohen's $d = 2.42$; 9-year-olds, 93%, $t(19) = 23.56$, $p < 0.001$, Cohen's $d = 5.27$; 11-year-olds, 88%, $t(19) = 7.37$, $p < 0.001$, Cohen's $d = 1.65$.

To directly compare space-pitch comprehension in Dutch and Turkish, children's binary responses on each trial were analyzed using a mixed effects logistic regression model. Analyses were performed in R (R version 4.0.4) using the afex package (Singmann et al., 2021) and the lmerTest package (Kuznetsova et al., 2015). Model terms were tested with likelihood ratio tests using the afex package (Singmann et al., 2021). Significant interactions were followed-up using Tukey contrasts to control for multiple comparisons using the emmeans package (Lenth et al., 2021). To establish if non-significant comparisons were an indication that an effect was absent (or sufficiently small), we performed equivalence tests (Lakens, 2017; Lakens et al., 2018); details of the procedure are outlined in the Supporting Information S1. The model included language (Dutch vs. Turkish), age group (5, 7, 9, 11 years), and their interaction as fixed effects, and subject as a random effect. Deviation coding was used for all factors (i.e., language and age group). All figures were produced in JASP (JASP Team, 2021).

The analysis revealed a significant interaction between language and age group, $\chi^2 = 11.42$, $p = 0.010$, as well as significant main effects of language, $\chi^2 = 41.86$, $p < 0.001$, and age group, $\chi^2 = 41.86$, $p < 0.001$. Post-hoc contrasts revealed that 5-year-old, 7-year-old and 9-year-old Turkish speaking children comprehended thickness-pitch terminology better than same-aged Dutch children comprehended height-pitch terminology: 5-year-olds, z.ratio = 2.07, $p = 0.04$; 7-year-olds, z.ratio = 4.30, $p < 0.001$; 9-year-olds, z.ratio = 5.56, $p < 0.001$. There was no difference between 11-year-old Dutch and Turkish children in space-pitch comprehension of their native language terms, z.ratio = 1.48, $p = 0.14$, *ns* (see Figure 3). However, according to an equivalence test, the group differences cannot be interpreted as statistically equivalent, z.ratio = −0.15, $p = 0.44$.

3.1.2 | Children's understanding of non-native pitch vocabulary

Next, we analyzed children's comprehension of space-pitch vocabulary not expressed in their native language (i.e., height-pitch terminology for Turkish; thickness-pitch for Dutch).

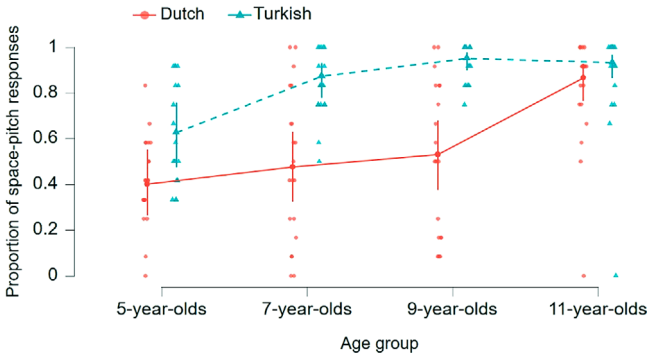


FIGURE 3 Dutch and Turkish children's comprehension of their native language space-pitch vocabulary. Data points of individual children are plotted (averaged across experimental trials). Group averages and standard deviations are also plotted for each age group

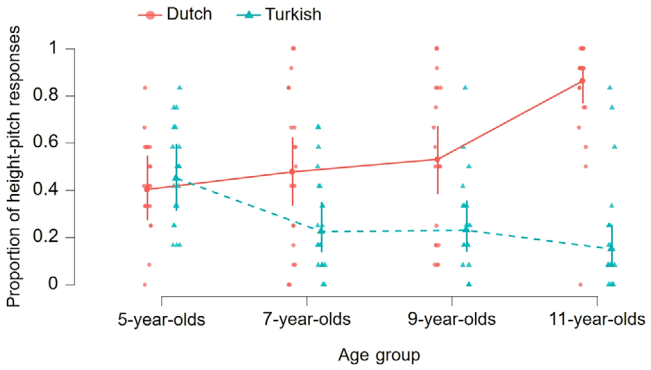


FIGURE 4 Dutch and Turkish children's height-pitch comprehension. Data points of individual children are plotted (averaged across experimental trials). Group averages and standard deviations are also plotted for each age group

Height-pitch vocabulary

Turkish children did not display better than chance comprehension of height-pitch language; in fact, they had on average 30% correct responses (collapsed across age groups). The youngest age group of 5-year-olds performed at chance with 46% correct responses, $t(19) = -0.89$, $p = 0.38$, Cohen's $d = -0.20$, but older Turkish speaking children were significantly below chance, thus reversing height-pitch terminology: 7-year-olds, 27%, $t(19) = -4.86$, $p < 0.001$, Cohen's $d = -1.09$; 9-year-olds, 27%, $t(19) = -5.07$, $p < 0.001$, Cohen's $d = -1.13$; 11-year-olds, 21%, $t(19) = -5.34$, $p < 0.001$, Cohen's $d = -1.20$.

To compare height-pitch comprehension in Turkish and Dutch children directly, performance on the height-pitch comprehension task was analyzed using a mixed effects logistic regression model with the same approach as previously, that is, language (Dutch vs. Turkish), age group (5, 7, 9, 11 years), and their interaction were fixed effects, and subject was included as a random effect. There was a significant language by age group interaction, $\chi^2 = 34.05$, $p < 0.001$, as well as a significant main effect of language, $\chi^2 = 39.92$, $p < 0.001$. The main effect of age group was not significant, $\chi^2 = 5.84$, $p = 0.12$, *ns*. Planned post-hoc contrasts (Tukey corrected) revealed 5-year-old Dutch children did

14677687, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/desc.13341 by Koc University, Wiley Online Library on [22/02/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

TABLE 2 Number of children reversing height-pitch vocabulary (i.e., scoring below 0.5) in the comprehension task (percentages in parentheses)

	5-year-olds	7-year-olds	9-year-olds	11-year-olds
Dutch	11 (55%)	10 (50%)	7 (35%)	1 (5%)
Turkish	10 (50%)	16 (80%)	17 (85%)	17 (85%)

not differ from 5-year-old Turkish speaking children, z.ratio = 0.46, $p = .65$, *ns*, although according to an equivalence test, the group differences cannot be interpreted as statistically equivalent, z.ratio = −1.56, $p = 0.06$. By 7 years, however, Dutch and Turkish children differed significantly in their understanding of height-pitch terminology with Turkish children more likely to reverse height-pitch terminology than Dutch children: 7-year-olds, z.ratio = −2.62, $p = 0.01$; 9-year-olds, z.ratio = −3.04, $p = 0.002$; and 11-year-olds, z.ratio = −7.65, $p < 0.001$ (see Figure 4).

To summarize, Turkish children were more likely to reverse height-pitch vocabulary than Dutch children. Intriguingly, around half the 5- and 7-year-old Dutch children also comprehended height-pitch terminology in a reversed fashion (see Table 2).

Thickness-pitch vocabulary

In contrast to Turkish children's behavior with height-pitch vocabulary, Dutch children did very well with thickness-pitch vocabulary, with 82% correct responses on average. Even the youngest children performed above chance: 5-year-olds, 75%, $t(19) = 5.74$, $p = < 0.001$, Cohen's $d = 1.28$; 7-year-olds, 83%, $t(19) = 9.41$, $p < 0.001$, Cohen's $d = 2.10$; 9-year-olds, 82%, $t(19) = 7.72$, $p < 0.001$, Cohen's $d = 1.73$ and 11-year-olds, 88%, $t(19) = 9.74$, $p < 0.001$, Cohen's $d = 2.18$.

Again, using a mixed effects logistic regression model, we found a significant interaction between language and age, $\chi^2 = 9.19$, $p < 0.03$, a significant main effect of age group, $\chi^2 = 30.91$, $p < 0.001$, but no main effect of language, $\chi^2 = 0.08$, $p = 0.77$, *ns*. Planned post-hoc contrasts (Tukey corrected) revealed 7- and 11-year-old Dutch children did not differ statistically from Turkish children: 7-year-olds, z.ratio = 0.03, $p = 0.98$, *ns*; 11-year-olds, z.ratio = 0.09, $p = 0.93$, *ns*. Furthermore, equivalence tests for 7-year-olds, z.ratio = 1.95, $p = 0.03$, and 11-year-olds, z.ratio = −1.68, $p = 0.05$ were significant, indicating 7- and 11-year-old Turkish and Dutch children's performance is statistically equivalent (i.e., close to zero). There were also significant differences at other ages, but these were not stable. Dutch children were significantly better than Turkish children at 5 years, z.ratio = −2.10, $p = 0.04$, but worse at 9 years, z.ratio = 2.21, $p = 0.03$ (see Figure 5).

3.2 | Nonlinguistic associations between space and pitch

We turn now to the non-linguistic tasks to establish whether non-linguistic space-pitch mappings differ between Dutch and Turkish children.

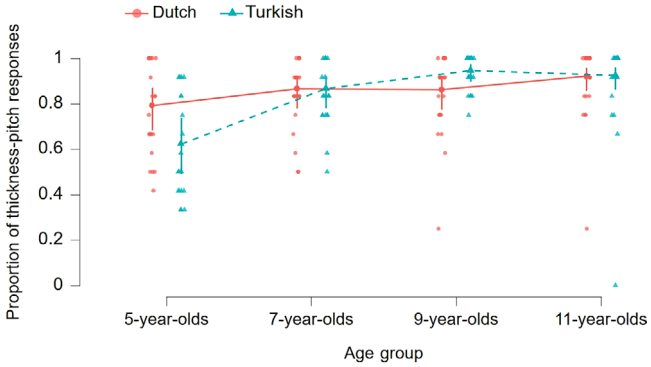


FIGURE 5 Dutch and Turkish children's thickness-pitch comprehension. Average performance and standard deviations are plotted for each age group. Data points display individual children (collapsed across experimental trials).

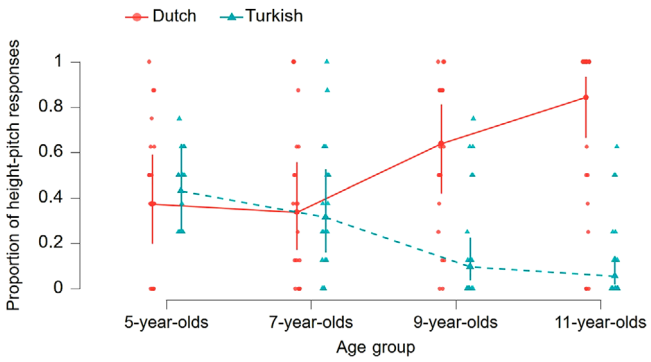


FIGURE 6 Dutch and Turkish children's nonlinguistic height-pitch associations. Data points of individual children are plotted (averaged across experimental trials). Group averages and standard deviations are also plotted for each age group.

3.2.1 | Nonlinguistic height-pitch associations

We first examined height-pitch associations and found performance overall was low: Dutch children had on average 53% correct, whereas Turkish children had 28% correct. As before, we conducted one-sample t-tests against chance (50%) for each language and age group. Dutch children were not able to match tones to spatial height until 11 years, 71%, $t(19) = 2.42$, $p = 0.03$, Cohen's $d = 0.54$. At younger ages Dutch children were at chance: 5-year-olds, 43%, $t(19) = -1.07$, $p = 0.30$, Cohen's $d = -0.24$; 7-year-olds, 39%, $t(19) = -1.32$, $p = 0.20$, Cohen's $d = -0.29$; and 9-year-olds, 59%, $t(19) = 1.20$, $p = 0.24$, Cohen's $d = 0.27$.

Turkish children showed a different profile. Although the youngest groups performed at chance; 5-year-olds, 44%, $t(19) = -1.53$, $p = 0.14$, Cohen's $d = -0.34$; 7-year-olds, 37%, $t(19) = -2.06$, $p = 0.05$, Cohen's $d = -0.46$, by 9 years Turkish children performed significantly below chance, 19%, $t(19) = -5.38$, $p < 0.001$, Cohen's $d = -1.20$; 11-year-olds, 13%, $t(19) = -8.39$, $p < 0.001$, Cohen's $d = -1.88$ (see Figure 6).

To summarize, older Turkish children consistently reversed height-pitch mappings by associating a low-pitched tone with a stimulus

14677687, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/desc.13341 by Koc University, Wiley Online Library on [22/02/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

TABLE 3 Number of children reversing height-pitch mappings (i.e. scoring below 0.5) in the non-linguistic task (percentages in parentheses)

	5-year-olds	7-year-olds	9-year-olds	11-year-olds
Dutch	11 (55%)	12 (60%)	6 (30%)	5 (25%)
Turkish	8 (40%)	12 (60%)	15 (75%)	17 (85%)

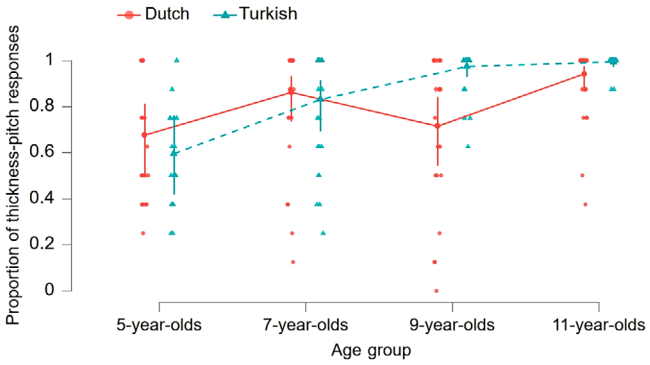


FIGURE 7 Dutch and Turkish children's nonlinguistic thickness-pitch associations. Data points of individual children are plotted (averaged across experimental trials). Group averages and standard deviations are also plotted for each age group.

located high in space and vice versa for low-pitched sounds. This is visible in individual level data too, as seen in Table 3.

3.2.2 | Nonlinguistic thickness-pitch associations

Unlike height-pitch associations, children performed well on thickness-pitch associations, with 74% correct responses for Dutch children and 82% correct for Turkish children. Even at the youngest ages, Dutch children were able to do this mapping better than chance, with a slight blip at 9 years where the threshold for statistical significance was not met: 5-year-olds, 63%, $t(19) = 2.15$, $p = 0.04$, Cohen's $d = 0.48$; 7-year-olds, 78%, $t(19) = 4.44$, $p < 0.001$, Cohen's $d = 0.99$; 9-year-olds, 66%, $t(19) = 2.14$, $p = 0.05$, Cohen's $d = 0.48$; and 11-year-olds, 89%, $t(19) = 9.58$, $p < 0.001$, Cohen's $d = 0.80$.

Turkish children showed a slight delay in the non-linguistic thickness-pitch mapping, with 5-year-olds at chance, 58%, $t(19) = 1.64$, $p = 0.12$, Cohen's $d = 0.37$, but by 7 years there was clear evidence of the association: 7-year-olds, 76%, $t(19) = 4.50$, $p < 0.001$, Cohen's $d = 1.01$; 9-year-olds, 94%, $t(19) = 17.90$, $p < 0.001$, Cohen's $d = 4.00$; 11-year-olds, 99%, $t(19) = 56.66$, $p < 0.001$, Cohen's $d = 12.67$ (see Figure 7).

3.3 | Comparing Dutch and Turkish children in height-pitch and thickness-pitch tasks

To assess whether non-linguistic space-pitch associations differed in language-specific ways, we analyzed responses using a mixed effects

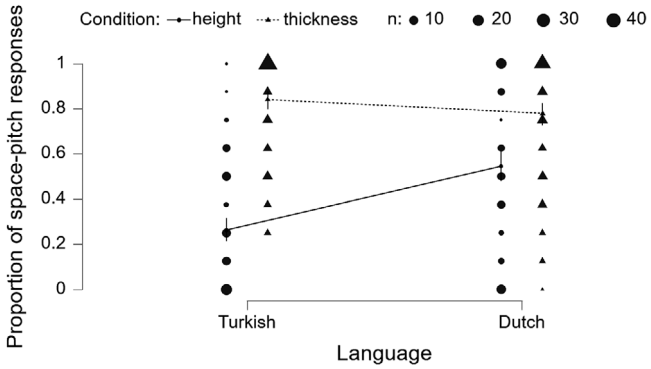


FIGURE 8 Dutch and Turkish children's nonlinguistic space-pitch associations (collapsed across age groups). Average performance and standard deviations are plotted. Data points display the performance of individual children (averaged across experimental trials), with the size of the data points corresponding to the number of children who yielded this performance

logistic regression model, including language (Dutch vs. Turkish), condition (height-pitch vs. thickness-pitch), age group (5, 7, 9, 11 years) and their interaction as fixed effects, and subject as a random effect. However, this full model failed to converge, so we collapsed across age groups by removing age group. This analysis revealed significant effects of condition, $\chi^2 = 412.49$, $p < 0.001$, language, $\chi^2 = 5.59$, $p = 0.02$, and an interaction between the two, $\chi^2 = 69.78$, $p < 0.001$. Post-hoc contrasts (Tukey corrected) revealed Dutch and Turkish children differed in their nonlinguistic height-pitch associations, z.ratio = −6.32, $p < 0.001$, but only marginally in their thickness-pitch associations, z.ratio = 1.96, $p = 0.05$ (see Figure 8). An equivalence test, however, revealed Dutch and Turkish children's thickness-pitch associations were statistically equivalent (i.e., close to zero), z.ratio = −2.21, $p = 0.01$.

To examine when children's nonlinguistic mappings started to diverge, we examined the effects of language (Dutch vs. Turkish) and condition (height-pitch vs. thickness-pitch) separately for each age group. At the youngest ages we did not find evidence that language influenced responses, but there was a difference across conditions. Both Dutch and Turkish children were better able to map thickness to pitch than height to pitch according to their canonical mappings. This was evident in the logistic regression models. For 5-year-olds, there was a significant main effect of condition, $\chi^2 = 18.55$, $p < 0.001$, but no significant effect of language, $\chi^2 = 0.11$, $p = 0.74$, and no interaction between condition and language, $\chi^2 = 0.88$, $p = 0.25$. The same was true for 7-year-olds: there was a significant main effect of condition, $\chi^2 = 114.78$, $p < 0.001$, but not of language, $\chi^2 = 0.22$, $p = 0.64$, and no interaction between condition and language, $\chi^2 = 0.05$, $p = 0.82$.

By 9 years the picture looked different. Now, as well as a significant main effect of condition, $\chi^2 = 143.98$, $p < 0.001$, there was a significant interaction between condition and language, $\chi^2 = 106.59$, $p < 0.001$, although no main effect of language, $\chi^2 = 0.07$, $p = 0.79$. Post-hoc contrasts (Tukey corrected) confirmed that at this age Dutch and Turkish children differed in both nonlinguistic height-pitch associations, z.ratio = 5.24, $p < 0.001$, and thickness-pitch associations,

14677687, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/desc.13341 by Koc University, Wiley Online Library on [22/02/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

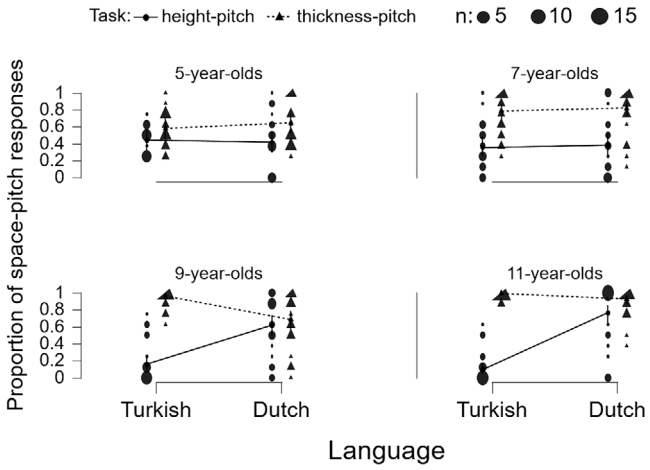


FIGURE 9 Dutch and Turkish children's performance on the nonlinguistic space-pitch association tasks, plotted separately for the different age groups. Average performance and standard deviations are plotted. Data points display the performance of individual children (averaged across experimental trials), with the size of the data points corresponding to the number of children who yielded this performance.

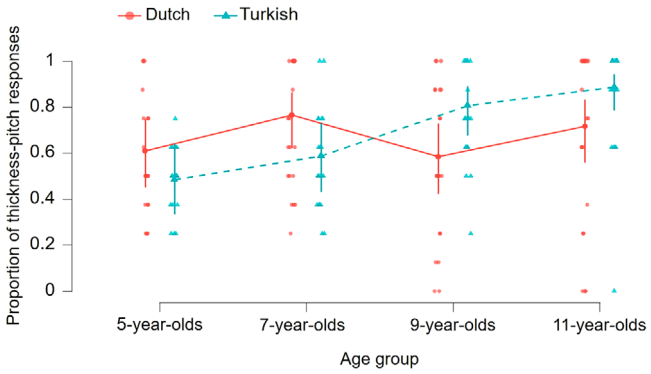


FIGURE 10 Dutch and Turkish children's performance on the nonlinguistic space-pitch conflict task. Average performance and standard deviations are plotted for each age group. Data points display the performance of individual children (collapsed across experimental trials).

z.ratio = 4.74, $p < 0.001$. In line with patterns in language, 9-year-old Dutch children performed better than Turkish children on the height-pitch association task, but Turkish children performed better on the thickness-pitch association task.

The same pattern was evident for 11-year-olds. There was no main effect of language, $\chi^2 = 1.66$, $p = 0.20$, but a significant main effect of condition, $\chi^2 = 242.78$, $p < 0.001$, and a significant interaction between condition and language, $\chi^2 = 79.72$, $p < 0.001$.[2] Post-hoc contrasts (Tukey corrected) revealed 11-year-old Dutch children performed better than same-aged Turkish children on the height-pitch association task, z.ratio = −7.56, $p < 0.001$, but Turkish children outperformed Dutch children in the thickness-pitch association task, z.ratio = 2.68, $p = 0.04$ (see Figure 9).

3.4 | Pitting height-pitch and thickness-pitch against each other

In the final task, children matched auditory pitch to spatial stimuli juxtaposing thickness and height (see Figure 2c). Overall, children opted for thickness over height, regardless of language background. Children's responses were analyzed using a mixed effects logistic regression model, including language (Dutch vs. Turkish), and age group (5, 7, 9, 11 years), and interaction as fixed effects, and subject as a random effect, and thickness-pitch responses on each trial as the dependent variable (note, a thickness response implies the rejection of a height response, and vice versa). While there was no significant main effect of language, $\chi^2 = 0.86$, $p = 0.36$, *ns*, there was a significant main effect of age group, $\chi^2 = 14.12$, $p = 0.003$, as well as a significant language by age group interaction, $\chi^2 = 13.37$, $p = 0.004$. Post-hoc contrasts (Tukey corrected) revealed no significant differences between younger Dutch children and same-aged Turkish speaking children, 5-year-olds, z.ratio = −1.10, $p = 0.27$, *ns*; 7-year-olds, z.ratio = −1.73, $p = 0.08$, *ns*. However, these differences were not statistically equivalent, 5-year-olds, z.ratio = −0.67, $p = 0.25$, *ns*; 7-year-olds, z.ratio = 0.02, $p = 0.51$, *ns*. By contrast, older Turkish children more often opted for a thickness-pitch interpretation than Dutch children of the same age, 9-year-olds, z.ratio = 2.25, $p = 0.03$, and 11-year-olds, z.ratio = 2.21, $p = 0.03$ (see Figure 10 and Supporting Information S1 for further analyses S1).

4 | GENERAL DISCUSSION

Overall, this study points to commonalities at early ages in children's knowledge of one of the building blocks of music, that is, pitch, but also demonstrates divergence across cultures by middle childhood, consistent with the proposal that language plays a formative role in space-pitch associations.

4.1 | Children's acquisition of pitch terminology

Dutch children are exposed to height-pitch vocabulary, whereas Turkish children are exposed to thickness-pitch vocabulary, giving rise to different learning conditions for children of each language. We found Turkish children acquire thickness-pitch terminology earlier than Dutch children acquire height-pitch terminology. While Turkish children had mastered thickness-pitch terminology from 7 years, only the oldest Dutch children (i.e., 11-year-olds) showed reliable height-pitch comprehension. These findings cannot be attributed to the polysemy of pitch vocabularies alone since both height-pitch and thickness-pitch terms have dual spatial and auditory meanings.

Although children have difficulties acquiring pitch terminology (e.g., Costa-Giomi & Descombes, 1996; Durkin & Townsend, 1997; Sergeant, 1984), our findings suggest polysemous expressions are not difficult per se. Rather, children's difficulties depend on the specific

14677687, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/desc.13341 by Koc University, Wiley Online Library on [22/02/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

spatial dimension used for expressing pitch. Thickness-pitch terminology appears easier to grasp than height-pitch terminology, as is also evident from the comprehension of novel space-pitch expressions (i.e., height-pitch for Turkish kids and thickness-pitch for Dutch kids). Even though Dutch does not express pitch in terms of spatial thickness, all Dutch children were above chance in their comprehension of thickness-pitch terminology. Dutch children's abilities were comparable to Turkish children, and even exceeded Turkish children at 5 years of age. However, later—by 9 years—Turkish children showed better thickness-pitch comprehension, likely scaffolded by additional input from language.

The fact that children, regardless of language background, showed excellent comprehension of thickness-pitch terminology, suggests this is more intuitive for the child learner. Although pitch is considered a metathetic dimension with no clear assignment of what constitutes 'more' or 'less' (Stevens, 1957), children may initially interpret low frequencies as 'more' and high frequencies as 'less' in pitch, perhaps because pitch and loudness are integral dimensions (Grau & Nelson, 1988; Melara & Marks, 1990) and loudness can be conceptualized as more or less. Similar mappings have been observed in early development for the mapping of achromatic color (i.e., lightness), where dark grey is initially perceived as 'more' than light grey (Smith & Sera, 1992).

Associating low (or high) pitch with more (or less) could also explain children's difficulties with height-pitch terminology, especially when it comes to the reversal of height-pitch mappings. We found Dutch children did not reliably understand height-pitch terminology until 11 years, and at earlier ages showed the reverse interpretation of height-pitch terminology (see also Dolscheid et al., 2015; Sergeant, 1984 for converging evidence). It seems younger children intuitively link low pitch with the unmarked term 'high' and high pitch with 'low', in opposition with the mapping direction of their input language (where 'high' is used for high pitch, and vice versa for 'low'). Reversal of height-pitch terminology was also observed in Turkish children with most children reversing height-pitch terms. It is possible that Turkish children's reversal of height-pitch terminology comes about because they draw an analogy to thickness-pitch terminology, interpreting 'high' as analogous to 'thick' and thus to designate 'low in pitch'.

Overall, then, both Turkish and Dutch children showed a tendency to reverse height-pitch terminology, but this was more consistent and persistent (encompassing older age groups) for Turkish children. In comparison, Dutch speaking children were ultimately able to map spatial height and pitch in accordance with their native language. So, over a protracted learning period, children's comprehension of pitch terminology comes to align with the norms of their speech community.

4.2 | Cross-cultural stability of thickness-pitch mappings

We find remarkable consistency for thickness-pitch associations across age and culture. Children were able to make this association, regardless of language. Furthermore, children's mappings appear comparable to those of adults. In a previous study, Dolscheid et al. (2020) reported adult Turkish speakers made non-linguistic thickness-pitch mappings correctly on average 87%, while in the current study Turkish 9-year-olds were on average 94% correct. A direct comparison across studies is not possible because of differences in the number of trials and testing conditions, but this points to remarkable stability. Even when children were presented with conflicting information in the space-pitch conflict task, they were more likely to opt for a thickness-pitch interpretation (e.g., when presented with a low pitch sound, children chose a thick line despite the fact it was high in space). This was true even though differences in spatial thickness were subtler in the conflict task than in the thickness-pitch association task, again supporting the conclusion that thickness-pitch mappings are more intuitive than height-pitch associations.

This differential behavior in mapping across space-pitch mappings requires explanation. Perhaps associations between pitch and thickness (and correspondingly size) are more biologically relevant and this makes them more available (e.g., Pietraszewski et al., 2017). Differences in body-size covary with pitch in most mammals, with larger species producing lower frequency sounds than smaller species. In addition, pitch is reliably associated with smaller objects in the real world. For instance, smaller musical instruments make higher frequency sounds than larger musical instruments. Children could pick up these statistical regularities, resulting in strong associations between thickness (size) and pitch, irrespective of language and culture.

4.3 | The vulnerability of height-pitch mappings

Unlike thickness-pitch associations, we find height-pitch mappings are more fragile than previously assumed. Dutch children comprehend height-pitch associations only at the age of 11, whereas Turkish speaking children consistently reverse associations between spatial height and pitch. Earlier studies found associations between spatial height and pitch in young infants (e.g., Dolscheid et al., 2014; Walker et al., 2018), but our data suggests a protracted developmental process instead.

This, therefore, raises questions about how to reconcile the emerging conflicting data. We consider a number of possibilities. First, infant studies make use of indirect, implicit paradigms (such as preferential looking), whereas older children are typically asked for explicit judgments instead. Second, there are differences in experimental materials. Infant studies commonly make use of pitch glides (e.g., Dolscheid et al., 2014; Walker et al., 2010; 2013), but child studies—including our own—often employ punctual auditory stimuli instead (e.g., Shayan et al., 2014; Starr & Srinivasan, 2018; Starr et al., 2021). Since infants associate spatial height with pitch only when dynamic stimuli, such as pitch glides and moving objects are used (cf., Jeschonek et al., 2013), this suggests dynamic trajectories could serve as an additional boost for linking spatial height and pitch. Third, there appear to be discrepancies in the pitch ranges used across studies. Most infant studies employ a broad pitch range exceeding 1 kHz, but the same is not true of child studies. Intriguingly, when Parise et al. (2014) recorded natural sounds from the environment, they found a consistent mapping between the frequency of sounds and the elevation of their sources in external space

14677687, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/desc.13341 by Koc University, Wiley Online Library on [22/02/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

that was particularly pronounced in the middle range of the spectrum (between 1 and 6 kHz) but seemed weaker—and even reversed—below 1 kHz (Parise et al., 2014). So, employing stimuli from the low end of the pitch spectrum may result in weaker or even reversed height-pitch mappings.

However, none of these factors explain inconsistencies between our findings and previous studies showing height-pitch associations can be mastered by young children (e.g., Starr & Srinivasan, 2018; Starr et al., 2021). Our study differs from Starr & Srinivasan (2018) in a number of ways, including differences in the ages tested (5-to-11-year-olds vs. 3-to-6-year-olds) as well as experimental materials (e.g., abstract forms, i.e., circles and squares vs. cartoon-like characters). Critically, even though Starr and Srinivasan report above chance associations between spatial height and auditory pitch in children, these do not seem entirely robust: when collapsed across age groups, average performance was only at 58%. In contrast, Dolscheid et al. (2020) report a non-linguistic height-pitch association of 92% for Dutch adults, suggesting the height-pitch mappings reported by Starr and Srinivasan are far from adult-like. Taken together, our findings suggest height-pitch associations undergo substantial developmental change (see also Speed et al., 2021).

4.4 | The role of language in space-pitch associations

Given the developmental and cross-cultural patterns in how children link space and pitch, it appears the role of language varies for different space-pitch associations. Language does not seem necessary to form thickness-pitch associations, since Dutch children show evidence of this association without being acquainted with thickness-pitch terminology in language. This is in contrast to the findings of Shayan et al. (2014) where only those 2- to 5-year-old children exposed to thickness-pitch terminology formed corresponding thickness-pitch associations. But other studies—like ours—have found children can make thickness-pitch mappings without the supporting input of language (e.g. Starr & Srinivasan, 2018; Starr et al., 2021).

Although language does not appear to be necessary for establishing thickness-pitch mappings, we propose it plays a foundational role in developing height-pitch associations. We found 11-year-old Dutch children who had mastered height-pitch terminology also reliably associated spatial height to pitch, whereas other cohorts performed at chance or even reversed height-pitch associations. However, since our data are correlational, they do not license conclusions about the directionality of causation. Future studies could target this issue directly; in the meantime our study provides evidence consistent with this interpretation relying on a cross-sectional approach.

Taken together, our findings point to diverging developmental pathways for different types of space-pitch mappings. Thickness-pitch associations appear stable but can be additionally enhanced by language, whereas height-pitch associations appear more fragile and susceptible to linguistic influence. In line with this, nonlinguistic height-pitch mappings are significantly weaker in adults when they are not additionally bolstered by language—as demonstrated by adult speakers of Catalan and Spanish (Fernandez-Prieto et al., 2017), Farsi (Holler et al., 2022), and Turkish (Dolscheid et al., 2020)—in comparison to speakers of height-pitch languages like English or Dutch (e.g., Dolscheid et al., 2020; Fernandez-Prieto et al., 2017).

4.5 | Implications for music development

Our study speaks to several open issues concerning the development of music cognition. By taking a cross-cultural and developmental approach, we show both commonalities and differences in how children acquire musical properties, such as musical pitch. We show previously attested difficulties in labeling pitch cannot be attributed to polysemy alone. Instead it is the specific alignment between spatial height and pitch that is difficult for the child learner. Thickness-pitch terminology is also polysemous, but is more intuitive and understood earlier by children. These findings can be informative for music education. To overcome children's difficulties with height-pitch vocabulary, for example, one could combine low-pitch tones with spatial stimuli that are both low in space and big in size (and vice versa for high-pitched sounds). In other words, one could take advantage of children's size-pitch associations for reinforcing height-pitch mappings.

Beyond music education, our findings also shed new light on children's nonlinguistic pitch representations. Although young infants and neonates are sensitive to associations between spatial dimensions and pitch (Dolscheid et al., 2014; Walker et al., 2018), adults are affected by differences in the pitch vocabularies they use (e.g., Dolscheid et al., 2013, 2020; Fernandez-Prieto et al., 2017). Here we add an important missing piece, linking the work from infants to adults. By considering the development of children into middle childhood, we reveal a more nuanced picture of how representations of pitch emerge and change. We find remarkable similarities in how children from different cultures map spatial thickness to pitch, suggesting statistical regularities resurface in the way pitch gets represented across cultures. This points to cross-cultural commonalities—and potentially even universal tendencies—akin to what has been observed in music (e.g. Mehr et al., 2019) or the perception of odor pleasantness (Arshamian et al., 2022). At the same time, our findings highlight cross-cultural differences in child music cognition. From 9 years of age, we find language-specific differences in how children link spatial dimensions and pitch, in line with the proposal by Lucy (2016) that linguistic relativity effects only appear around middle childhood, after 8-years of age.

Our findings highlight the importance of cross-cultural empirical studies to disentangle what is cross-culturally stable versus variable in how children come to represent musical pitch, similar to what has been observed for other dimensions of music (e.g., Athanasopoulos et al., 2021; Jacoby et al., 2019; McPherson et al., 2020). Children's pitch representations ultimately become attuned to patterns in language, but only later in development, underlining the fact that the most protracted processes in music cognition are those that are the most culture specific (Stalinski & Schellenberg, 2012).

14677687, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/desc.13341 by Koc University, Wiley Online Library on [22/02/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

5 | CONCLUSIONS

Children from different cultures show differences in their understanding of music vocabulary and in their nonlinguistic associations between spatial dimensions and auditory pitch. While height-pitch mappings are acquired late and seem to require additional scaffolding from language, thickness-pitch mappings are acquired early and are less susceptible to language. In sum, our study provides new evidence that space-pitch mappings change over development, showing signatures of both cross-cultural stability and diversity.

ORCID

Sarah Dolscheid https://orcid.org/0000-0001-9104-6386

NOTES

[1] There were no transgender or gender nonbinary children in this sample, so *n* minus *female* provides the number of male children tested.

[2] It has to be noted that—despite the simple structure of the random effects—the model failed to converge. However, a logistic regression (without random effects) yielded the same results.

ACKNOWLEDGMENTS

The authors would like to thank all of the children, their families, and schools for participating in this research. Thanks to Ilona Plug for help with data collection. This research was supported by an Ammodo Science Award to Asifa Majid.

CONFLICT OF INTEREST

None.

DATA AVAILABILITY STATEMENT

The data are publicly available at OSF: https://osf.io/vgfzr/

REFERENCES

Andrews, M. L., & Madeira, S. S. (1977). The assessment of pitch discrimination ability in young children. *Journal of Speech and Hearing Disorders*, *42*(2), 279–287. https://doi.org/10.1044/jshd.4202.279

Arshamian, A., Gerkin, R. C., Kruspe, N., Wnuk, E., Floyd, S., O'Meara, C., Garrido Rodriguez, G., Lundström, J. N., Mainland, J. D., & Majid, A. (2022). The perception of odor pleasantness is shared across cultures. *Current Biology*, *32*(9), 2061–2066.e3. https://doi.org/10.1016/j.cub.2022.02.062

Athanasopoulos, G., Eerola, T., Lahdelma, I., Kaliakatsos-Papakostas, M., & Damjanovic, L. (2021). Harmonic organisation conveys both universal and culture-specific cues for emotional expression in music. *Plos One*, *16*(1), 1–17. https://doi.org/10.1371/journal.pone.0244964

Braginsky, M., Yurovsky, D., Marchman, V. A., & Frank, M. C. (2019). Consistency and variability in children's word learning across languages. *Open Mind*, *3*, 52–67. https://doi.org/10.1162/opmi_a_00026

Costa-Giomi, E., & Descombes, V. (1996). Pitch labels with single and multiple meanings: A study with French-speaking children. *Journal of Research in Music Education*, *44*(3), 204–214. https://doi.org/10.2307/3345594

Dolscheid, S., Çelik, S., Erkan, H., Küntay, A., & Majid, A. (2020). Space-pitch associations differ in their susceptibility to language. *Cognition*, *196*(September 2019), 104073. https://doi.org/10.1016/j.cognition.2019.104073

Dolscheid, S., Hunnius, S., Casasanto, D., & Majid, A. (2014). Prelinguistic infants are sensitive to space-pitch associations found across cultures. *Psychological Science*, *25*(6), 1256–1261. https://doi.org/10.1177/0956797614528521

Dolscheid, S., Hunnius, S., & Majid, A. (2015). When high pitches sound low: Children's acquisition of space-pitch metaphors. In (N. Miyake, D. Peebles, & R. P. Cooper, Eds.), *Proceedings of the 34th Annual Meeting of the Cognitive Science Society* (pp. 306–311). Cognitive Science Society.

Dolscheid, S., Shayan, S., Majid, A., & Casasanto, D. (2013). The thickness of musical pitch: Psychophysical evidence for linguistic relativity. *Psychological Science*, *24*(5), 613–621. https://doi.org/10.1177/0956797612457374

Durkin, K., & Shire, B. (1990). Children's linguistic difficulties in a pitch directionality task involving cross-modal stimuli. *Educational Psychology*, *10*(2), 169–179. https://doi.org/10.1080/0144341900100204

Durkin, K., & Townsend, J. (1997). Influence of linguistic factors on young school children's responses to musical pitch tests: A preliminary test. *Psychology of Music*, *25*(2), 186–191. https://doi.org/10.1177/0305735697252007

Eitan, Z., & Timmers, R. (2009). Beethoven's last piano sonata and those who follow crocodiles: Cross-domain mappings of auditory pitch in a musical context. *Cognition*, *114*(3), 405–422. https://doi.org/10.1016/j.cognition.2009.10.013

Fernandez-Prieto, I., Spence, C., Pons, F., & Navarra, J. (2017). Does language influence the vertical representation of auditory pitch and loudness? *I-Perception*, *8*(3), 1–11. https://doi.org/10.1177/2041669517716183

Grau, J. W., & Nelson, D. K. (1988). The distinction between integral and separable dimensions: Evidence for the integrality of pitch and loudness. *Journal of Experimental Psychology: General*, *117*(4), 347–370. https://doi.org/10.1037/0096-3445.117.4.347

Hannon, E. E., Nederlanden, V. B. d., C, M., & Tichko, P. (2012). Effects of perceptual experience on children's and adults' perception of unfamiliar rhythms. *Annals of the New York Academy of Sciences*, *1252*(1), 92–99. https://doi.org/10.1111/j.1749-6632.2012.06466.x

Henrich, J., Heine, S. J., & Norenzayan, A. (2010). The weirdest people in the world? *Behavioral and Brain Sciences*, *33*(2–3), 61–83. https://doi.org/10.1017/S0140525X0999152X

Jacoby, N., Undurraga, E. A., McPherson, M. J., Valdés, J., Ossandón, T., & McDermott, J. H. (2019). Universal and non-universal features of musical pitch perception revealed by singing. *Current Biology*, *29*(19), 3229–3243.e12. https://doi.org/10.1016/j.cub.2019.08.020

JASP Team (2021). *Jasp (0.14.1)*.

Jeschonek, S., Pauen, S., & Babocsai, L. (2013). Cross-modal mapping of visual and acoustic displays in infants: The effect of dynamic and static components. *European Journal of Developmental Psychology*, *10*(3), 337–358. https://doi.org/10.1080/17405629.2012.681590

Kidd, E., & Garcia, R. (2022). How diverse is child language acquisition research? *First Language*, *42*(6), 703–735. https://doi.org/10.1177/01427237211066405

Kuznetsova, A., Brockhof, P. B., & Christensen, R. H. B. (2015). *lmerTest: Tests in linear mixed effects models. R package*. https://cran.r-project.org/web/packages/lmerTest/index.html

Lakens, D. (2017). Equivalence tests: A practical primer for t tests, correlations, and meta-analyses. *Social Psychological and Personality Science*, *8*(4), 355–362. https://doi.org/10.1177/1948550617697177

Lakens, D., Scheel, A. M., & Isager, P. M. (2018). Equivalence testing for psychological research: A tutorial. *Advances in Methods and Practices in Psychological Science*, *1*(2), 259–269. https://doi.org/10.1177/2515245918770963

Lenth, R. V., Buerkner, P., Herve, M., Love, J., Miguez, F., Riebl, H., & Singmann, H. (2021). *emmeans: Estimated marginal means, aka least-squares means. R package (version 1.5.4.)*. https://cran.r-project.org/package=emmeans

Lewkowicz, D. J., & Minar, N. J. (2014). Infants are not sensitive to synesthetic cross-modality correspondences: A comment on walker et al. (2010). *Psychological Science*, *25*(3), 832–834. https://doi.org/10.1177/0956797613516011

14677687, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/desc.13341 by Koc University, Wiley Online Library on [22/02/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

Lucy, J. A. (2016). The implications of linguistic relativity for language learning. In (R. Alonso Alonso, Ed.), *Crosslinguistic influence in second language acquisition*. ,(Vol. Vol. 15, (Issue 2), pp. 1–23).

Majid, A., Roberts, S. G., Cilissen, L., Emmorey, K., Nicodemus, B., O'Grady, L., Woll, B., LeLan, B., de Sousa, H., Cansler, B. L., Shayan, S., de Vos, C., Senft, G., Enfield, N. J., Razak, R. A., Fedden, S., Tufvesson, S., Dingemanse, M., Ozturk, O., … & Levinson, S. C. (2018). Differential coding of perception in the world's languages. *Proceedings of the National Academy of Sciences*, *115*(45), 11369–11376. https://doi.org/10.1073/pnas.1720419115

McPherson, M. J., Dolan, S. E., Durango, A., Ossandon, T., Valdés, J., Undurraga, E. A., Jacoby, N., Godoy, R. A., & McDermott, J. H. (2020). Perceptual fusion of musical notes by native amazonians suggests universal representations of musical intervals. *Nature Communications*, *11*(1), 1–14. https://doi.org/10.1038/s41467-020-16448-6

Mehr, S. A., Singh, M., Knox, D., Ketter, D. M., Pickens-Jones, D., Atwood, S., Lucas, C., Jacoby, N., Egner, A. A., Hopkins, E. J., Howard, R. M., Hartshorne, J. K., Jennings, M. V., Simson, J., Bainbridge, C. M., Pinker, S., O'Donnell, T. J., Krasnow, M. M., & Glowacki, L. (2019). Universality and diversity in human song. *Science*, *366*(6468), https://doi.org/10.1126/science.aax0868

Melara, R. D., & Marks, L. E. (1990). Perceptual primacy of dimensions: Support for a model of dimensional interaction. *Journal of Experimental Psychology: Human Perception and Performance*, *16*(2), 398–414..

Nettl, B. (2000). An ethnomusicologist contemplates universals in musical sound and musical culture. In (N. L. Wallin, B. Merker, & S. Brown, Eds.), *The origins of music*. (pp. 463–472). MIT Press.

Parise, C. V., Knorre, K., & Ernst, M. O. (2014). Natural auditory scene statistics shapes human spatial hearing. *Proceedings of the National Academy of Sciences*, *111*(16), 6104–6108. https://doi.org/10.1073/pnas.1322705111

Pietraszewski, D., Wertz, A. E., Bryant, G. A., & Wynn, K. (2017). Three-month-old human infants use vocal cues of body size. *Proceedings of the Royal Society B: Biological Sciences*, *284*(1856), 20170656. https://doi.org/10.1098/rspb.2017.0656

Pratt, C. C. (1930). The spatial character of high and low tones. *Journal of Experimental Psychology*, *13*(3), 278–285. https://doi.org/10.1037/h0072651

Sergeant, D. (1984). A language for auditory space. *Early Child Development and Care*, *14*(1–2), 37–74. https://doi.org/10.1080/0300443840140103

Shayan, S., Ozturk, O., Bowerman, M., & Majid, A. (2014). Spatial metaphor in language can promote the development of cross-modal mappings in children. *Developmental Science*, *17*(4), 636–643. https://doi.org/10.1111/desc.12157

Shayan, S., Öztürk, Ö., & Sicoli, M. (2011). The thickness of pitch: Crossmodal metaphors in farsi, turkish, and zapotec. *The Senses and Society*, *6*(1), 96–105. https://doi.org/10.2752/174589311%D7;12893982233911

Singmann, H., Bolker, B., Westfall, J., Aust, F., & Ben-Shachar, M. S. (2021). *afex: Analysis of factorial experiments. R package* (version 0.28-1). https://cran.r-project.org/package=afex

Smith, L. B., & Sera, M. D. (1992). A developmental analysis of the polar structure of dimensions. *Cognitive Psychology*, *24*(1), 99–142. https://doi.org/10.1016/0010-0285(92)90004-L

Speed, L. J., Croijmans, I., Dolscheid, S., & Majid, A. (2021). Crossmodal associations with olfactory, auditory, and tactile stimuli in children and adults. *I-Perception*, *12*(6), https://doi.org/10.1177/20416695211048513

Stalinski, S. M., & Schellenberg, E. G. (2012). Music cognition: A developmental perspective. *Topics in Cognitive Science*, *4*(4), 485–497. https://doi.org/10.1111/j.1756-8765.2012.01217.x

Stalinski, S. M., Schellenberg, E. G., & Trehub, S. E. (2008). Developmental changes in the perception of pitch contour: Distinguishing up from down. *The Journal of the Acoustical Society of America*, *124*(3), 1759–1763. https://doi.org/10.1121/1.2956470

Starr, A., Cirolia, A. J., Tillman, K. A., & Srinivasan, M. (2021). Spatial metaphor facilitates word learning. *Child Development*, *92*(3), e329–e342. https://doi.org/10.1111/cdev.13477

Starr, A., & Srinivasan, M. (2018). Spatial metaphor and the development of cross-domain mappings in early childhood. *Developmental Psychology*, *54*(10), 1822–1832. https://doi.org/10.1037/dev0000573

Stevens, S. S. (1957). On the psychophysical law. *The Psychological Review*, *64*(3), 153–181. https://doi.org/10.1126/science.3.71.712.b

Trehub, S. E. (2000). Human processing predispositions and musical universals. In (N. L. Wallin, M. Merker, & S. Brown, Eds.), *The origins of music*. (pp. 427–448). MIT Press.

Trehub, S. E., Becker, J., & Morley, I. (2015). Cross-cultural perspectives on music and musicality. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *370*(1664). https://doi.org/10.1098/rstb.2014.0096

Trehub, S. E., Schellenberg, G. E., & Nakata, T. (2008). Cross-cultural perspectives on pitch memory. *Journal of Experimental Child Psychology*, *100*(1), 40–52. https://doi.org/10.1016/j.jecp.2008.01.007

Walker, P., Bremner, J. G., Lunghi, M., Dolscheid, S. D., Barba, B., & Simion, F. (2018). Newborns are sensitive to the correspondence between auditory pitch and visuospatial elevation. *Developmental Psychobiology*, *60*(2), 216–223. https://doi.org/10.1002/dev.21603

Walker, P., Bremner, J. G., Mason, U., Spring, J., Mattock, K., Slater, A., & Johnson, S. P. (2010). Preverbal infants' sensitivity to synaesthetic cross-modality correspondences. *Psychological Science*, *21*(1), 21–25. https://doi.org/10.1177/0956797609354734

Webster, P. R., & Schlentrich, K. (1982). Discrimination of pitch direction by preschool children with verbal and nonverbal tasks. *Journal of Research in Music Education*, *30*(3), 151–161. https://doi.org/10.2307/3345082

Weiss, M. W., Cirelli, L. K., Mcdermott, J. H., & Trehub, S. E. (2020). Development of consonance preferences in western listeners. *Journal of Experimental Psychology: General*, *149*(4), 634–649. https://doi.org/10.1037/xge0000680

Zemp, H., & Malkus, V. (1979). Aspects of 'Are'are musical theory. *Ethnomusicology*, *23*(1), 5–48. https://doi.org/10.2307/851336

SUPPORTING INFORMATION

Additional supporting information can be found online in the Supporting Information section at the end of this article.

How to cite this article: Dolscheid, S., Çelik, S., Erkan, H., Küntay, A., & Majid, A. (2022). Children's associations between space and pitch are differentially shaped by language. *Developmental Science*, e13341. https://doi.org/10.1111/desc.13341

14677687, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/desc.13341 by Koc University, Wiley Online Library on [22/02/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License