Brief article

Space-pitch associations differ in their susceptibility to language

Sarah Dolscheid[a], Simge Çelik[b], Hasan Erkan[c], Aylin Küntay[d], Asifa Majid[c,e,f,*]

[a] *University of Cologne, Cologne, Germany*
[b] *Heidelberg University, Heidelberg, Germany*
[c] *Radboud University, Nijmegen, the Netherlands*
[d] *Koç University, Istanbul, Turkey*
[e] *Max Planck Institute for Psycholinguistics, Nijmegen, the Netherlands*
[f] *University of York, Heslington, UK*

ARTICLE INFO

ABSTRACT

To what extent are links between musical pitch and space universal, and to what extent are they shaped by language? There is contradictory evidence in support of both universality and linguistic relativity presently, leaving the question open. To address this, speakers of Dutch who talk about pitch in terms of spatial height and speakers of Turkish who use a thickness metaphor were tested in simple nonlinguistic space-pitch association tasks. Both groups showed evidence of a thickness-pitch association, but differed significantly in their height-pitch associations, suggesting the latter may be more susceptible to language. When participants had to match pitches to spatial stimuli where height and thickness were opposed (i.e., a thick line high in space vs. a thin line low in space), Dutch and Turkish differed in their relative preferences. Whereas Turkish participants predominantly opted for a thickness-pitch interpretation—even if this meant a reversal of height-pitch mappings—Dutch participants favored a height-pitch interpretation more often. These findings provide new evidence that speakers of different languages vary in their space-pitch associations, while at the same time showing such associations are not equally susceptible to linguistic influences. Some space-pitch (i.e., height-pitch) associations are more malleable than others (i.e., thickness-pitch).

1. Introduction

Musical pitch is intricately linked to representations of space (Möhring, Ramsook, Hirsh-Pasek, Golinkoff, & Newcombe, 2016; Pratt, 1930; Rusconi, Kwan, Giordano, Umiltà, & Butterworth, 2006). Participants are faster responding to higher tones when a button is located higher in space, while the opposite is true for lower tones (Rusconi et al., 2006). People also link spatial size and pitch, associating higher pitches with smaller objects and lower pitches with larger objects (e.g., Parise & Spence, 2009). Even infants and newborns seem to be sensitive to associations between some spatial dimensions and pitch (e.g., Dolscheid, Hunnius, Casasanto, & Majid, 2014; Pietraszewski, Wertz, Bryant, & Wynn, 2017; Walker et al., 2010, 2018; but see Lewkowicz & Minar et al., 2014).

At the same time, space-pitch associations can be shaped by musical experience (Taylor & Witt, 2015), as well as metaphors in language (e.g., Dolscheid, Shayan, Majid, & Casasanto, 2013; Fernandez-Prieto, Spence, Pons, & Navarra, 2017). Languages like English, Dutch, and German express musical pitch in terms of spatial height, for example,

but Farsi speakers use a thickness-metaphor instead, with 'thin' referring to higher pitches and 'thick' to lower pitches (Shayan, Öztürk, & Sicoli, 2011). This cross-linguistic variation is also reflected in concomitant differences in thinking about sound (Dolscheid et al., 2013). Whereas Dutch speakers' pitch representations are influenced by spatial height—but not thickness—in space-pitch interference tasks, the reverse holds true for speakers of Farsi who show sensitivity to thickness but not height, supporting the notion that language plays a critical role in space-pitch associations (Dolscheid et al., 2013).

Other research challenges this conclusion, however. For instance, German adults reliably associate thickness and pitch in a simple association task despite the absence of a thickness-pitch metaphor in German (Shayan, Ozturk, Bowerman, & Majid, 2014), and so do English speaking children (Starr & Srinivasan, 2018). Furthermore, although Farsi participants' pitch estimates were not affected by spatial height (Dolscheid et al., 2013), the Kreung in northeastern Cambodia—who are reported not to use height-pitch metaphors in their language—apparently exhibit height-pitch associations (Parkinson, Kohler, Sievers, & Wheatley, 2012).

* Corresponding author at: University of York, Heslington, YO10 5DD, UK
E-mail address: asifa.majid@york.ac.uk (A. Majid).

These conflicting findings warrant further cross-linguistic examination. We therefore tested speakers of Dutch who use a height-pitch metaphor and compared them to speakers of a new language in this context—Turkish—who like Farsi speakers use a thickness-pitch metaphor (Majid et al., 2018; Shayan et al., 2011). Both groups were tested in simple non-linguistic height-pitch and thickness-pitch association tasks. Given the previous findings, it is possible that Dutch and Turkish speakers have both height- and thickness-pitch mappings available to them. If so, it is also conceivable that the mappings are not equally salient to both groups—a possibility that has not been explored previously. Height-pitch associations reveal themselves in a variety of tasks (Chiou & Rich, 2012; Maeda, Kanai, & Shimojo, 2014; Rusconi et al., 2006), are present from as early as 44 hours from birth (Walker et al., 2018), and have been said to be universally present in languages (Evans & Treisman, 2010; Pratt, 1930; Stumpf, 1883). It is claimed this mapping reflects real-world auditory scene statistics—i.e., higher frequency sounds come from higher in space—and that the convoluted anatomy of the outer ear evolved to mirror these auditory statistics (Parise, Knorre, & Ernst, 2014). Taken together, this suggests height-pitch associations should be privileged.

To test this, we introduced a novel paradigm designed to establish the relative importance of the two space-pitch mappings directly. We made participants choose between a height or thickness spatial association to a sound by presenting them with a single tone and asking them to choose either a *thick* line presented *high* in space or a *thin* line presented *low* in space (see Fig. 1a). This paradigm forces participants to choose only one mapping, thereby elucidating which of the two may be stronger. Our predictions were the following: if space-pitch associations are universal, then there should be comparable associations for the two groups across tasks. Alternatively, if the language a person speaks influences space-pitch mappings, Dutch and Turkish speakers' associations should differ corresponding to patterns in language.

2. Methods

2.1. Participants

Forty native speakers of Turkish (age: $M = 21.3$, $SD = 2.01$, 20 female) and 40 native speakers of Dutch (age: $M = 21.8$, $SD = 2.79$, 32 female) were paid for participation. Dutch speakers were recruited and tested in Nijmegen, the Netherlands, and Turkish speakers in Istanbul, Turkey. We also tested a different cohort of 24 native Turkish speakers in an online experiment (age: $M = 27.9$, $SD = 3.43$, 12 female).

2.2. Materials and procedure

Participants completed a task where they matched pitches to one of two stimuli with conflicting spatial information (space-pitch conflict task). Subsequently, participants were tested in a height-pitch association task and a thickness-pitch association task, with the order of presentation counterbalanced between participants. To avoid spill-over effects from the two space-pitch association tasks, the conflict task was always administered first. Participants were tested individually in a quiet room. The tasks were presented on a PC laptop using Presentation software (version 18.1, www.neurobs.com). Sounds were presented via headphones at approximately 60 dB-a. The volume was kept constant across participants. Instructions were given in the participants' native language with no space-pitch metaphors used during testing. An online version of the height-pitch association task was administered separately using OpenSesame (Mathôt, Schreij, & Theeuwes, 2012) and JATOS (Lange, Kühn, & Filevich, 2015).

2.3. Nonlinguistic space-pitch conflict task

In the conflict task, participants first listened to one of two pure tones for 2 seconds (698 vs. 440 Hz). They then saw two horizontal lines of equal length (6.9 cm) presented side-by-side on the screen. One line was relatively thick (9 mm width) and was presented at the top of a 28.5 x 18.5 cm grid of white dots on a black background; the other line was relatively thin (1.5 mm width) and presented at the bottom of the grid (see Fig. 1a). Participants were asked to choose which of the two visual stimuli best matched the sound by pressing the corresponding button (i.e., left vs. right) on a Cedrus RB540 response box. They were asked to make their responses as quickly and accurately as possible. Stimuli disappeared as soon as a button was pressed. If a decision was not made within a 30 seconds window while the image was displayed, participants were asked to make a decision afterwards when the image disappeared. Before the actual experiment, two practice trials were given in order to familiarize participants with the task. For the experimental trials, the location of the spatial stimuli (left vs. right) was fully crossed with the two tones (high vs. low) and repeated 8 times. The resulting 32 experimental trials were presented randomly and no feedback was given. Responses were coded as reflecting either a height-pitch mapping or a thickness-pitch mapping.

2.4. Nonlinguistic space-pitch association tasks

The space-pitch association tasks were similar to the conflict task except for details of the spatial stimuli and number of trials. In the height-pitch task, participants were presented with two orange balls (both approximately 2.2 cm diameter), presented side-by-side, with one ball at the top and the other at the bottom of a 12.5 x 11.3 cm grid of small white dots on a black background (see Fig. 1b). The same task was administered in the online version of the study.

In the thickness-pitch task two vertical tubes of orange color and equal length (11.5 cm) were presented side-by-side, with one tube being
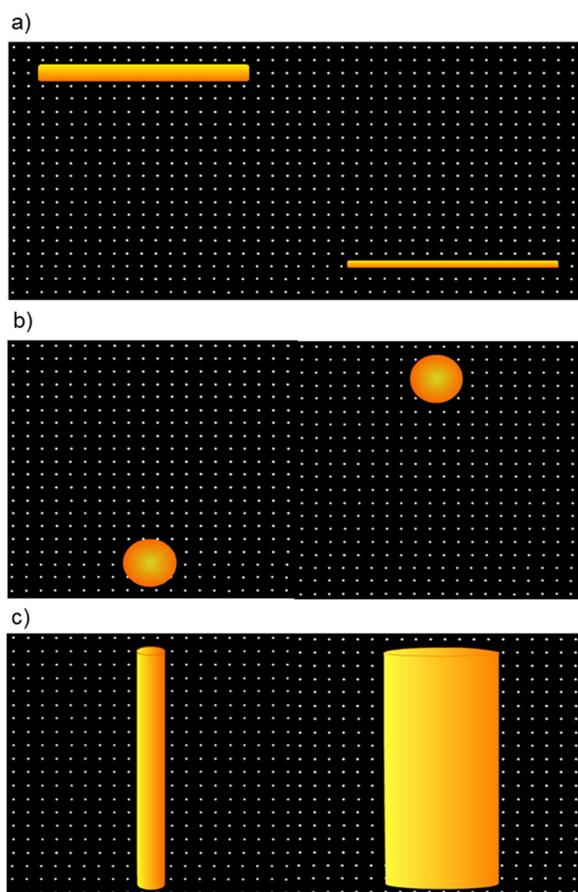

Fig. 1. Example visual stimuli for the: (a) space-pitch conflict task, (b) height-pitch association task, and (c) thickness-pitch association task.

rather thin (1.2 cm width) and the other comparatively thick (4.7 cm width, see Fig. 1b). For both tasks there were 8 trials.

2.5. Music reading ability

To control for differences in musical skills (in particular participants' ability to read musical notes), participants also filled out a background questionnaire where they noted their music reading skills on a 7-point Likert-scale (from $1 = not at all$ to $7 = fluently$).

3. Results

3.1. Music reading ability

In our sample, Turkish speaking participants reported better music reading abilities ($M = 3.53$; $SD = 1.92$) than Dutch participants ($M = 2.08$; $SD = 1.99$), $t(78) = 3.31$, $p = .001$, $d = .74$. Music reading for Turkish speakers tested in the online version of the height-pitch association task ($M = 2.75$; $SD = 1.67$) did not differ significantly from the original cohort of Turkish participants, $t(62) = 1.39$, $p = .17$, ns, $d = .36$, nor from the Dutch speakers, $t(62) = 1.64$, $p = .11$, ns, $d = .42$. Music reading ability was therefore included as a variable in the critical analyses to follow.

3.2. Nonlinguistic space-pitch conflict task

Analyses were carried out using R (R Core Team, 2013, version 3.4.2) with packages lme4 (Bates, Maechler, & Bolker, 2012) and languageR (Baayen, 2009). For the conflict task, we performed mixed effects logistic regression analyses of language (Turkish, Dutch) on participants' preference for height-pitch mappings (since participants decide between height and thickness mappings on a single trial, a height response implies the rejection of a thickness response, and vice versa). We also added music-reading ability as a predictor. Using the principle of backward selection, a full model was defined by including random intercepts for subjects and items as well as by-subject and by-item random slopes for the effect of music-reading (i.e., within-subject factor). Since the full model failed to converge, only random intercepts for subjects and items were included, with p-values computed by likelihood ratio tests. We found a significant main effect of language, $\chi^2(1) = 12.87$, $p < .001$, but no effect of music-reading on participants' height-pitch preference, $\chi^2(1) = .12$, $p = .73$, ns.

To establish whether participants' performance differed from chance, one-sample t-tests were computed for each group. Turkish speakers' proportion of height-pitch preference was significantly below chance, $M = 15\%$, $t(39) = -9.71$, $p < .001$, $d = -1.54$, indicating a thickness-pitch preference. On the other hand, Dutch speakers' proportion of height-responses did not differ from chance, $M = 50\%$, $t(39) = .06$, $p = .95$, $d = .01$. This outcome is compatible with two different scenarios: either Dutch speakers were split into two groups (i.e., those who preferred a height-pitch mapping vs. those who preferred a thickness-pitch mapping); or Dutch speakers did not have consistent space-pitch mappings in this task, but instead switched preferences between trials. To distinguish between these two possibilities, we identified those participants who did not show a clear response strategy (i.e., whose preference for height-pitch mappings was between 44% and 56%). This pattern applied to only 5 Dutch participants and 1 Turkish participant. When these participants were excluded, 36 of the remaining 39 Turkish speaking participants (92%) opted for thickness-pitch; whereas Dutch speakers were indeed split into two groups—-around half (17/35 = 49%) preferred height-pitch in line with their language's metaphors, whereas half (18/35 = 51%) preferred thickness-pitch (see Fig. 2).

3.3. Nonlinguistic space-pitch association tasks

Mixed effects logistic regression analyses were performed with language (Turkish, Dutch), task (height, thickness), as well as music-reading ability as predictors, and with model fitting as described previously. There was a main effect of language $\chi^2(1) = 52.17$, $p < .001$, and task $\chi^2(1) = 292.02$, $p < .001$, but no significant effect of music-reading $\chi^2(1) = 3.56$, $p = .06$. Crucially, the model also yielded a significant interaction of language by task $\chi^2(1) = 67.73$, $p < .001$, such that height-pitch associations differed between Turkish and Dutch participants, $\chi^2(1) = 54.72$, $p < .001$; but not thickness-pitch associations, $\chi^2(1) = .91$, $p = .34$. As expected on the basis of linguistic metaphors, Turkish speakers showed stronger associations between thickness and pitch than height and pitch, $\chi^2(1) = 353.85$, $p < .001$; but Dutch speakers' performance was equivalent for height-pitch and thickness-pitch associations, $\chi^2(1) = 1.76$, $p = .18$ (see Fig. 3). To test the possibility that the observed differences were introduced by spill-over effects from the conflict task, we compared height-pitch associations from the new cohort of Turkish speakers who were tested only in the height-pitch association task (i.e., the online task) to the Dutch speakers. The new cohort of Turkish speakers also displayed significantly weaker height-pitch associations than Dutch speakers, $\chi^2(1) = 17.37$, $p < .001$.

To establish whether participants' performance differed from chance, one-sample t-tests were computed for each group and task. Dutch participants performed above chance for both thickness-pitch, $M = 95\%$, $t(39) = 16.30$, $p < .001$, $d = 2.58$ and height-pitch tasks, $M = 92\%$, $t(39) = 21.13$, $p < .001$, $d = 3.34$. In contrast, Turkish speakers' tested under the same conditions were above chance for thickness-pitch, $M = 87\%$, $t(39) = 8.51$, $p < .001$, $d = 1.35$, but significantly below chance for height-pitch, $M = 32\%$, $t(39) = -2.79$, $p < .01$, $d = -.44$. So speakers of Turkish reversed the height-pitch mapping by associating higher positions with lower pitches, and vice versa. When tested exclusively on the height-pitch task online, however—i.e., without being exposed to the conflict task—Turkish participants performed above chance for height-pitch associations, $M = 66\%$, $t(23) = 10.47$, $p < .001$, $d = 2.14$ (see Fig. 3), suggesting that the space-pitch conflict task led to the reversal of height-pitch associations.

3.4. General discussion

Speakers of Turkish and Dutch not only use different space-pitch metaphors in language, they also associate pitch differently across spatial dimensions in simple non-linguistic tasks. When tested in a novel paradigm in which spatial dimensions were juxtaposed, Turkish and Dutch speakers solved the task in different ways. Whereas Turkish speakers overwhelmingly opted for the thickness-pitch mapping, only half the Dutch speakers did—the other half favored the height-pitch mapping instead. Participants opted for one solution and stuck with it throughout the course of the experiment (as has been demonstrated in other experiments with similar format; e.g., Lin & Murphy, 2001). While future studies should move beyond binary response options to examine the exact parameters that determine participants' space-pitch associations, here we provide the first evidence of differential weighting of simultaneously presented spatial features by speakers of different languages.

Contrary to what might have been predicted (cf., Parise et al., 2014), we find associations between spatial height and pitch appear to be more malleable and more susceptible to effects of language. Although both Dutch and Turkish speakers reliably associated height and pitch, this mapping was significantly weaker in speakers of Turkish. This was true even when spill-over effects from other tasks were ruled out, suggesting that associations between spatial height and pitch are more stable when they receive additional support from metaphors in language.
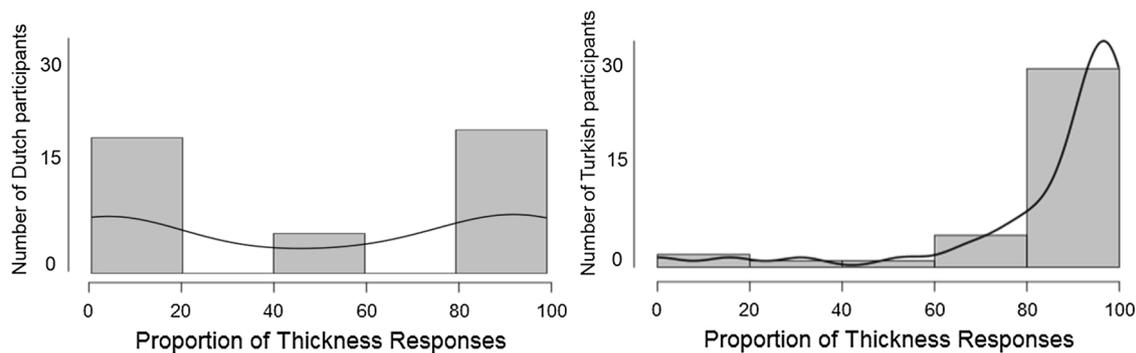
Fig. 2. Results of the space-pitch conflict task. Histograms of thickness-pitch associations for speakers of Dutch (left) and Turkish (right) with distribution curves of the datasets. Whereas Dutch speakers fall into two groups: those who predominantly opt for a height-pitch vs. a thickness-pitch mapping, the great majority of Turkish speakers favor the thickness-pitch association. Graphs generated in JASP (JASP Team, 2019).
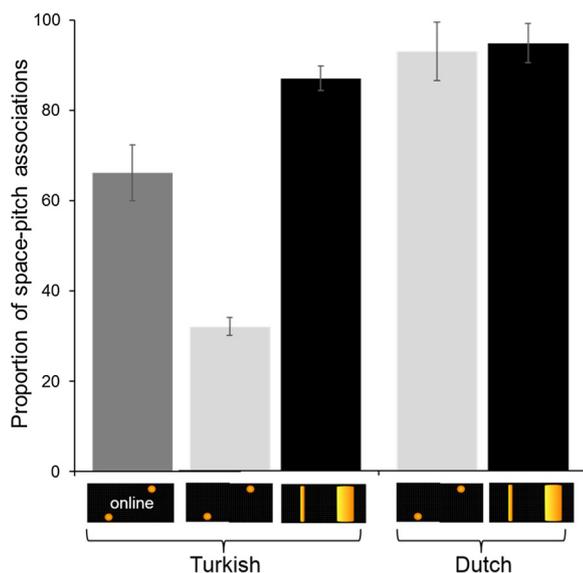

Fig. 3. Results of the non-linguistic space-pitch association tasks for Turkish and Dutch speakers. Canonical space-pitch associations (i.e., responses in line with height-pitch or thickness-pitch mappings) are depicted. Error bars represent standard error of the mean.

On the other hand, Dutch participants were able to associate spatial thickness and pitch, particularly in the simple thickness-pitch association task where they performed just like Turkish participants, even though they do not use thickness-pitch metaphors in their language. This suggests thickness-pitch associations may be less susceptible to linguistic effects. These findings are in line with previous demonstrations of thickness-pitch associations in the absence of corresponding linguistic metaphors (e.g., Shayan et al., 2011; Starr & Srinivasan, 2018). At the same time, other studies employing an interference paradigm with Dutch speakers did not find evidence of spatial thickness influencing pitch (Dolscheid et al., 2013), suggesting particulars of the task parameters are critical. Simple binary association tasks may be more likely to elicit consistent space-pitch mappings in comparison to other paradigms, such as the parametric space-pitch interference tasks of Dolscheid et al. (2013). Notably, however, Dolscheid and colleagues find that only a short amount of training is sufficient to produce thickness-pitch correspondences, suggesting they are nascent in the Dutch.

It is possible that associations between pitch and thickness (and thus size) are more stable since they are biologically relevant (e.g., Pietraszewski et al., 2017). Differences in body-size covary with pitch in most mammals, with smaller species producing higher frequency sounds than larger species. Cross-linguistic evidence also suggests that size-pitch—not height-pitch—associations are more frequently expressed across diverse languages (Majid et al., 2018), and linguistic thickness-pitch metaphors also seem more intuitive than height-pitch metaphors during language acquisition. For instance, Turkish speaking children acquire thickness-pitch linguistic metaphors earlier than German speaking children acquire height-pitch metaphors (Shayan et al., 2014).

Our findings highlight the importance of studying alternative space-pitch mappings by means of cross-linguistic studies. Focusing on mappings or languages in isolation (e.g., by only investigating a height-pitch mapping among the Kreung, see Parkinson et al., 2012) will likely underestimate existent cross-cultural variation. Rather than creating space-pitch mappings, language may come to affect the strength of a particular association relative to another. In line with this account, Fernandez-Prieto et al. (2017) found relatively stronger height-pitch mappings in speakers of English than Catalan or Spanish who make less consistent use of a height-pitch metaphor in language. Our findings support and extend these results by showing that height-pitch mappings can even be reversed in speakers of Turkish in certain circumstances: after exposure to the conflict task where spatial height and thickness information were in conflict, Turkish speakers inverted associations between height and pitch (i.e., by mapping high tones with low positions in space). This reversal was exclusively observed in speakers of Turkish—but not Dutch—supporting the proposal that height-pitch mappings are rather fragile when they are not bolstered by additional input from language.

The observed vulnerability of associations between spatial height and pitch seems at odds with the attested prevalence of height-pitch mappings across tasks and age-groups (e.g., Jeschonek, Pauen, & Babocsai, 2013; Pratt, 1930; Rusconi et al., 2006). Why this discrepancy? One possibility is that previous research has overestimated the robustness of height-pitch associations because it has focused almost exclusively on Western participants who are likely to be familiar with linguistic height-pitch metaphors. When considering a broader variety of languages, however, there are occasions where the height-pitch mapping is inverted. For instance, Farsi speakers who predominantly use a thickness-pitch metaphor in language, also use the term *boland* 'tall/high' to refer to low pitched tones (see Shayan et al., 2011), and show a tendency to reverse the height-pitch mapping in a non-linguistic height-pitch interference task (Dolscheid et al., 2013).

Another critical factor concerns the stimuli employed. For instance, Jeschonek et al. (2013) found infants associate spatial height with pitch, but this was only true when dynamic stimuli were used (i.e., pitch glides and moving objects). Here we presented participants with static stimuli, which arguably could have underestimated their height-pitch associations. However, since associations between height and pitch were not absent but weaker (or reversed) in Turkish speaking participants, the task we administered appears sufficiently sensitive to measure associations. While future studies could investigate the role of

stimulus properties more closely, here we provide new evidence for the role of language on space-pitch associations.

To conclude, we show that language can affect space-pitch mappings even in simple non-linguistic association tasks but that the influence varies across space-pitch mappings. Whereas thickness-pitch mappings appear comparable in Dutch and Turkish speakers, height-pitch associations differ, and can even be inverted in speakers of Turkish. When spatial dimensions are juxtaposed, Dutch and Turkish speaking participants interpreted the same stimuli in different ways, lending support to the idea that there are different relative weightings of thickness-pitch and height-pitch associations in the two groups. Our results suggest effects of language on space-pitch associations may be more complex than previously assumed. While mappings between space and pitch can be shaped by language, they also differ in their susceptibility to linguistic influences.

Declaration of Competing Interest

None.

Acknowledgements

References

Baayen, H. R. (2009). *LanguageR*. R Package.

Bates, D., Maechler, M., & Bolker, B. (2012). *lme4: Linear mixed-effects models using S4 classes*. R package version 0.999375-42. 2011Google Scholar.

Chiou, R., & Rich, A. N. (2012). Cross-modality correspondence between pitch and spatial location modulates attentional orienting. *Perception, 41*(3), 339–353. https://doi.org/10.1068/p7161.

Dolscheid, S., Çelik, S., Erkan, H., Küntay, A., & Majid, A. (2019). *Space-pitch associations differ in their susceptibility to language [datasets]*. [dataset] Retrieved from osf.io/csw9q. DOI 10.17605/OSF.IO/CSW9Q.

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

Evans, K. K., & Treisman, A. (2010). Natural cross-modal mappings between visual and auditory features. *Journal of Vision, 10*, 1–12. https://doi.org/10.1167/10.1.6. Introduction.

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), 204166951771618. https://doi.org/10.1177/2041669517716183.

JASP Team (2019). *JASP (Version 0.10.2)[Computer software]*.

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

Lange, K., Kühn, S., & Filevich, E. (2015). "Just another tool for online studies" (JATOS): An easy solution for setup and management of web servers supporting online studies. *PloS One, 10*(6), 1–14. https://doi.org/10.1371/journal.pone.0130834.

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.

Lin, E. L., & Murphy, G. L. (2001). Thematic relations in adults' concepts. *Journal of Experimental Psychology General, 130*(1), 3–28.

Maeda, F., Kanai, R., & Shimojo, S. (2014). Changing pitch induced visual motion illusion. *Current Biology, 14*(23), 990–991.

Majid, A., Roberts, S. G., Cilissen, L., Emmorey, K., Nicodemus, B., O'Grady, L., ... 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.

Mathôt, S., Schreij, D., & Theeuwes, J. (2012). OpenSesame: An open-source, graphical experiment builder for the social sciences. *Behavior Research Methods, 44*(2), 314–324. https://doi.org/10.3758/s13428-011-0168-7.

Möhring, W., Ramsook, K. A., Hirsh-Pasek, K., Golinkoff, R. M., & Newcombe, N. S. (2016). Where music meets space: Children's sensitivity to pitch intervals is related to their mental spatial transformation skills. *Cognition, 151*, 1–5. https://doi.org/10.1016/j.cognition.2016.02.016.

Parise, C. V., & Spence, C. (2009). 'When birds of a feather flock together': Synesthetic correspondences modulate audiovisual integration in non-synesthetes. *PloS One, 4*(5), 1–7. https://doi.org/10.1371/journal.pone.0005664.

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.

Parkinson, C., Kohler, P. J., Sievers, B., & Wheatley, T. (2012). Associations between auditory pitch and visual elevation do not depend on language: Evidence from a remote population. *Perception, 41*, 854–861. https://doi.org/10.1068/p7225.

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.

R Core Team (2013) R Foundation for Statistical Computing, Vienna, Austria. URL https://www.R-project.org/.

Rusconi, E., Kwan, B., Giordano, B. L., Umiltà, C., & Butterworth, B. (2006). Spatial representation of pitch height: The SMARC effect. *Cognition, 99*(2), 113–129. https://doi.org/10.1016/j.cognition.2005.01.004.

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. *The Senses and Society, 6*(1), 96–105.

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.

Stumpf, C. (1883). *Tonpsychologie*. Leipzig: S. Hirzel.

Taylor, J. E. T., & Witt, J. K. (2015). Listening to music primes space: Pianists, but not novices, simulate heard actions. *Psychological Research, 79*(2), 175–182. https://doi.org/10.1007/s00426-014-0544-x.

Walker, P., Bremner, J. G., Lunghi, M., Dolscheid, S., Barba, B. D., & 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.