The Sound of Thickness: Prelinguistic Infants' Associations of Space and Pitch

Sarah Dolscheid[1,2] (sarah.dolscheid@mpi.nl) **Sabine Hunnius[3]** (s.hunnius@donders.ru.nl) **Daniel Casasanto[4]** (casasand@newschool.edu) **Asifa Majid[1,3]** (asifa.majid@mpi.nl)

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

Abstract

People often talk about musical pitch in terms of spatial metaphors. In English, for instance, pitches can be *high* or *low*, whereas in other languages pitches are described as *thick* or *thin*. According to psychophysical studies, metaphors in language can also shape people's nonlinguistic space-pitch representations. But does language establish mappings between space and pitch in the first place or does it modify preexisting associations? Here we tested 4-month-old Dutch infants' sensitivity to height-pitch and thickness-pitch mappings in two preferential looking tasks. Dutch infants looked significantly longer at cross-modally congruent stimuli in both experiments, indicating that infants are sensitive to space-pitch associations prior to language. This early presence of space-pitch mappings suggests that these associations do not originate from language. Rather, language may build upon pre-existing mappings and change them gradually via some form of competitive associative learning.

Keywords: cross-modal; multisensory; metaphors; synaesthesia; infant perception; language acquisition; language of perception; preferential looking

Introduction

Does a cake taste yellow? Or a tone played by a trumpet sound scarlet? For some people they do. Yet synaesthesia, a condition in which stimulation of one sensory modality induces systematic perceptual experiences in another modality, is relatively rare. Other types of cross-modal associations, however, can be found in non-synaesthetes too. Psychophysical studies have shown that adults and children without synaesthesia associate higher pitches with sharper edges (Marks, 1987; Parise & Spence, 2009), positions higher in space (e.g., Ben-Artzi & Marks, 1995; Melara & O'Brien, 1987), lighter color (Hubbard 1996; Marks 1989; Melara 1989), and increasing brightness (Marks, 1987).

Even infants seem to be sensitive to some of these associations. Cross-modal correspondences between loudness and brightness have been demonstrated in 20- to 30-day-old infants (Lewkowicz & Turkewitz, 1980). In a preferential looking task, 3- to 4-month-old infants preferred congruent trials – in which visuospatial height and pitch height corresponded – over incongruent trials. (Walker et al., 2010); that is, infants looked longer at a ball moving upwards if it was accompanied by a rising pitch than if it was accompanied by a falling pitch. Also, prelinguistic infants under 1-year-old 'matched' visual arrows pointing up or down with tones sweeping up or down in frequency (Wagner et al., 1981).

These findings have led to the assumption that cross-modal mappings are innately hardwired in the brain (Mondloch & Maurer, 2004) and represent an unlearned aspect of perception (Walker et al., 2010). Accordingly, some of these associations are posited to be universal (Marks, Hammeal, & Bornstein, 1987; see also Spence, 2011).

However, there are other findings that seem to be at odds with these conclusions. A number of cross-modal correspondences are only acquired later in the course of development. For instance, even 9-year-old children are not able to systematically match size and pitch, a task that is consistently solved by adults (Marks et al., 1987; but see Mondloch & Maurer, 2004). Cross-modal correspondences are also affected by developmental changes. Unlike adults and older children, 2-year-olds consistently map light grey to smaller objects and dark grey to bigger objects (Smith & Sera, 1992). In the course of language acquisition, however, children's associations gradually shift. As a result, it has been suggested that language may have an impact on the trajectory of cross-modal relations (Smith & Sera, 1992).

On one hand, language appears to mirror cross-modal experience. The auditory domain, for instance, is often linguistically encoded in terms of other sensory modalities (Williams, 1976). People use metaphors like "soft" voice, "dark" timbre or "high" pitch, suggesting that language echoes cross-modal perceptual impressions. On the other hand, language also seems to affect cross-modal associations. For example, Martino & Marks (1999) suggest that cross-modal effects, like the association between space and pitch, may be mediated by language. Various tasks show correspondences between spatial height and pitch consistent with high-low metaphors in language (Rusconi, et al, 2006, Evans & Treisman, 2010). However, since linguistic labels and height-pitch associations merely coincide, the direction of influence is hard to establish and the contribution of language remains unclear.

Cross-linguistic comparison provides one way to overcome this limitation. Not every language uses the same metaphors for pitch (Ashley, 2004; Levinson & Majid, 2007). For example, while languages like English and Dutch talk about pitch in terms of "height", other languages like, Farsi, Turkish and Zapotec (spoken in Mexico) describe high-frequency pitch as "thin" and low-frequency pitches as "thick" (Shayan, Ozturk, & Sicoli, 2011). To find out whether these differences in spoken metaphors correspond to different mental representations of pitch, Dolscheid and colleagues, conducted a series of nonlinguistic psychophysical experiments in adult speakers of Dutch (a "height" language) and Farsi (a "thickness" language). Participants were asked to reproduce musical pitches that they heard in the presence of irrelevant spatial information, i.e. lines varying either in height or in thickness (Dolscheid et al., submitted). Dutch speakers' pitch estimates were significantly modulated by spatial height but not by thickness. Conversely, Farsi speakers' pitch estimates were modulated by spatial thickness but not by height. Overall, the results indicated that nonlinguistic pitch-space associations follow language-specific vocabulary, suggesting that cross-modal pitch representations are language-specific.

At this point, however, it is unclear whether language establishes cross-modal mappings between space and pitch in the first place, or whether it merely modifies preexisting associations. While some researchers stress the relevance of language in concept formation (e.g. Gopnik & Meltzoff, 1997), others argue that conceptual representations must precede the acquisition of language (e.g. Bloom, 2000; Bloom & Keil, 2001). The former position allows for a stronger role of language in space-pitch associations; while the latter position suggests that children are likely to have some notion of space-pitch correspondences prior to learning language. Consistent with this latter view, infants seem to be sensitive to height-pitch mappings even prelinguistically (Walker et al., 2010). Critically, however, nothing is known about the origins of thickness-pitch relationships. Do children also have this mapping available to them prior to language, or is it only learned on the basis of language-input?

One possibility is that children could start out with both a height-pitch and a thickness-pitch mapping even before they learn language. The strength of these mappings might then subsequently be adjusted, according to the relative frequencies of space-pitch metaphors in the languages children acquire (Casasanto 2008, 2010; Dolscheid et al., submitted; Smith & Sera, 1992). Alternatively, height-pitch and thickness-pitch associations might follow different trajectories. Whereas height-pitch associations are available to prelinguistic infants, the thickness-pitch mapping might only be learned later. Metaphors in language could provide one possible way to learn this association. Using thickness terminology to refer to pitch may invite speakers to align correspondent representations and extract similarities between space and pitch in a process called structural alignment (see e.g. Boroditsky, 2001; Gentner, 2003). In line with this proposal, Shayan et al. (submitted) found that Turkish and Farsi 2- to 5-year-olds were able to successfully map thickness to pitch but same-aged German children (who like English and Dutch speakers do not have a thickness metaphor) were not able to make this association successfully. This is consistent with the proposal that language input promotes cross-modal associations between thickness and pitch. Note, however, that these results do not rule out the possibility that the thickness-pitch mappings were available to all infants prelinguistically, but are no longer equally available to German children.

In order to determine the prelinguistic availability of space-pitch mappings, we tested 4-month-old Dutch babies using a preferential looking paradigm. To investigate height-pitch correspondences, we followed Walker et al.'s (2010) procedure. Infants watched a ball moving up and down the screen accompanied by the sound of a sliding whistle. The whistle's fundamental frequency changed at a constant rate. In the congruent condition, the pitch of the sound "rose" and "fell" in accordance with the movement of the ball. In the incongruent condition, the pitch of the sound "rose" and "fell" in opposition to the movement of the ball (see Fig. 1a). Walker et al. (2010) reported that infants looked longer at the congruent compared to the incongruent condition, suggesting an early preference for pitch-height congruencies.

In a second step, we tested prelinguistic infants in a thickness-pitch task analogous to the height-pitch task. Instead of balls moving up and down the screen, a vertical tube varied in thickness, changing continuously from thin to thick (see Fig. 1b).

We reasoned that if both, height-pitch and thickness-pitch mappings are available to infants prelinguistically, infants should prefer both congruent height-pitch and congruent thickness-pitch stimuli over incongruent ones. If however, height-pitch and thickness-pitch relationships follow different developmental trajectories, with thickness mappings only becoming acquired later, then infants should show preferences for congruent height-pitch stimuli but not for congruent thickness-pitch stimuli.

a

b

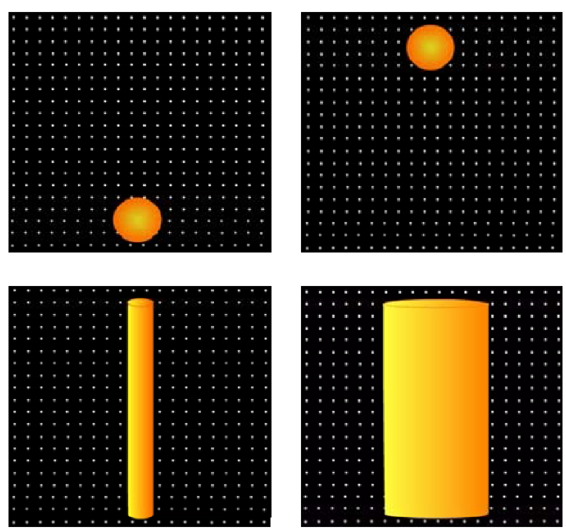

Figure 1: Examples of animations presented as stimuli in Experiment 1 (Panel a) and Experiment 2 (Panel b). In (a) the extremes of the ball's vertical trajectory are shown. In (b) the extremes of thickness are depicted. The images are reproduced to scale.

Experiment 1: Auditory pitch and visuospatial height

Methods

Participants Ten male and ten female infants completed the first (pitch-height) experiment (mean age = 129 days, range: 113 to 138 days). Another seven infants were tested, but not included in the analyses: one infant was excluded due to experimenter error; a further six infants were excluded due to excessive fussiness.

Materials and Procedure QuickTime animations were presented on a 102 x 76 cm Sony LCD screen using HABIT software. Animations appeared within a 67 x 67 cm screen area (25.6° x 25.6°), and lasted a maximum of 60 s. Before each animation, a flashing light appeared to ensure that infants attended to the screen. Infants sat in a Maxi-Cosi infant seat which was placed on their parent's lap, viewing the animations from a distance of 1.50 m. Infants' visual fixations were monitored and recorded on video. Animations were stopped if the infant looked elsewhere for a single period of 1 s or more. The total duration an infant looked directly at the animation was logged online during the experiment using HABIT software and written to an output-file. Additionally, looking times were determined by a subsequent frame-by-frame coding of the digitized video using SuperCoder. Coding was performed by a coder blind to the experimental condition. 25 percent of the data was double-coded by a second person, also blind to the condition.

Infants watched a 10-cm (4°) diameter orange ball moving up and down a 50-cm vertical trajectory in front of a 20 x 20 grid of small, white dots on a black field. The ball moved at a constant speed of 20 cm/s and paused for 42 ms at each endpoint. Animations were accompanied by the sound of a sliding whistle (a sinusoidal tone). The fundamental frequency of the sound changed at a constant rate, between 300 and 1700 Hz over 2.5 s, coinciding with a single phase of the animation (e.g. the ball moving up). The amplitude of the sound increased and then decreased between 47 and 84 dB within each phase of the animation, peaking at 1000 Hz. Amplitude thus changed about twice as fast as pitch to ensure that variation in perceived pitch and loudness were not confounded.

Every infant viewed three congruent animations interleaved with three incongruent animations. Half of the children watched a congruent animation first and the other half watched an incongruent animation first. During the whole experiment, parents were listening to music via headphones. Since parents could not hear the sliding sounds, they were unable to distinguish between experimental conditions (the spatial trajectories of the stimuli did not differ between conditions). We therefore ensured parents could not bias their infant's looking behavior.

Results

A high level of agreement was confirmed between the two observers in their coding of each infant's individual looking times (mean Pearson's $r(28) = .99$, $p<.001$).

14 of the 20 infants looked longer at the congruent animation than at the incongruent animation. On average, infants looked at the congruent animations for 31.7 s (*SD* = 11,4) and at the incongruent animations for 26.1 s (*SD* = 13.3). A paired-samples t-test confirmed that infants looked significantly longer at the congruent animations, $t(19) = 1.99$, $p = .03$, $d = 0.45$ (one-tailed).

Experiment 2: Auditory pitch and visuospatial thickness

Methods

Participants Ten male and ten female infants completed the second (pitch-thickness) experiment (mean age = 127 days, range: 113 to 138 days). An additional eight infants were tested, but their data was not analyzed: one infant was excluded due to technical problems; a further 7 infants were excluded due to fussiness.

Materials and Procedure The same procedure as in Experiment 1 was used. This time, infants watched a vertical orange tube that varied in thickness, changing continuously from thin to thick (see Figure 1).

The animation was presented on a 20 x 20 grid of small, white dots on a black field, as in Experiment 1. The tube was 60 cm long ranging from 6 to 26 cm in width. It expanded at a constant speed of 8 cm/s and paused for 42 ms at each endpoint. Animations were accompanied by the sound of the exact same sliding whistle as in Experiment 1. The fundamental frequency of the sound changed at a constant rate, between 300 and 1700 Hz over 2.5 s, coinciding with a single phase of the animation (i.e. during tube expansion). Each infant viewed three congruent animations interleaved with three incongruent animations, with half of the children watching a congruent animation first and the other half watching an incongruent animation first. During the whole experiment, parents were listening to music via headphones.

Results

A high level of agreement was confirmed between the two judges in their estimates of each infant's individual looking times (mean Pearson's $r(28) = .99$, $p<.001$).

13 of the 20 infants looked longer at the congruent animations than the incongruent animations. On average, infants looked longer at the congruent 24.4 s (*SD* = 11.8) than the incongruent animations 19.4 s (*SD* = 11.5). A paired-samples t-test confirmed that infants looked significantly longer at the congruent animations, $t(19) = 2.19$, $p = .02$, $d = 0.43$ (one-tailed).

Between experiment comparison

Dutch infants looked significantly longer at cross-modally congruent stimuli in both experiments. While this suggests a comparable starting point for both thickness-pitch and height-pitch mappings, it is nevertheless possible that infants display differential preference with respect to the two mappings. We therefore compared the results of the two previous experiments directly.

Results

Submitting looking times to a 2 (Space: height vs. thickness) by 2 (Congruency: congruent vs. incongruent) mixed ANOVA yielded a significant main effect of Space ($F(2,38) = 4.40$, $p = .04$, $\eta_p^2 = 0.10$), showing that looking times differed between height and thickness stimuli. Infants looked longer at height stimuli as compared to thickness stimuli, indicating that perhaps height was more salient for the them. However, no interaction between Space and Congruency ($F(2,38) = 0.03$, *ns*, $\eta_p^2 = 0.00$) was observed. There were thus no indications that looking time reductions induced by incongruency differed between the two experiments. In line with this, percentage reduction in looking time across experiments was of comparable size, i.e., 18% for the height-pitch experiment and 20% for the thickness-pitch experiment.

General Discussion

Our results demonstrate that prelinguistic infants are sensitive to correspondences between auditory pitch and spatial information of two different types, visuospatial thickness as well as height. Dutch infants looked significantly longer at cross-modally congruent stimuli in both experiments, suggesting a comparable starting point for height-pitch mappings and thickness-pitch mappings.

It is possible that these mappings are only present in very young infants but get lost in the course of development due to neuronal pruning. Whereas 2-to 3-month-old infants were found to be sensitive to arbitrary associations between colors and shapes, 8-month-old infants no longer show this early synaesthetic association (Wagner & Dobkins, 2011). Does the same developmental trajectory hold true for space-pitch mappings? Unlike synaesthetic color-shape associations that seem highly individualized and thus unspecific (e.g., one infant might associate triangles with green, and another with red), space-pitch associations follow a specific pattern, showing the same congruity preferences found in languages. It is therefore possible that space-pitch mappings persist during infancy and childhood. In line with this suggestion, sensitivity to height-pitch associations has been reported in 6-month-olds (Braaten, 1993) as well as in children aged 4 to 5 years (Roffler & Butler, 1967). On the other hand, 2- to 5-year-old German speaking children have been found to be insensitive to the thickness mapping (Shayan et al., submitted). There is also contradictory evidence regarding children's sensitivity to size-pitch associations. Whereas Marks et al. (1987) report that children are unable to systematically map size (big vs. small) to pitch until they are 13 years old, Mondloch and Maurer (2004) find evidence for size-pitch congruency effects in children as young as 3 years of age.[1] Details about the developmental trajectory of space-pitch mappings thus remain unclear and are subject to future research.

One aspect that seems to facilitate detecting cross-modal associations is motion (see also Jeschonek, Pauen & Babocsai, in preparation). Mondloch and Maurer presented children with moving balls that differed in size; while Marks et al. and Shayan et al. used static stimuli. In the present study, too, the dynamic display of spatial information (up- and downward movement or horizontal expansion) and pitch (presented as glides) may have facilitated the detection of corresponding information. Displaying stimuli dynamically and in synchronicity could direct infants' attention to the relational correspondences across modalities. However, movement by itself cannot explain the pattern of results: infants must still align stimuli attributes that are congruent to each other.

Language acquisition and cross-modal associations

Our findings demonstrate that both height-pitch and thickness-pitch correspondences are perceived before the infant has mastery of language. While this finding is consistent with the view that representations precede language (e.g. Bloom, 2000), it does not entail that these associations are fixed. Language could still influence the structure and content of preexisting mental representations via simple learning mechanisms. In the course of language acquisition, the relative strengths of different space-pitch mappings could be adjusted according to the language-specific frequencies of metaphors that children acquire (Casasanto 2008, 2010). Over time, speakers of a "height" language like Dutch would strengthen the height-pitch mapping at the expense of the thickness-pitch mapping – and vice versa for speakers of a "thickness" language like Farsi (Dolscheid et al., submitted). Evidence in support of this associative learning account is provided by linguistic training experiments. Dutch speakers, after being trained to use Farsi-like metaphors describing pitch relationships in terms of thickness, demonstrated nonlinguistic thickness-pitch mappings just like Farsi speakers. By contrast, when participants received the same amount of linguistic training with an alternative space-pitch mapping that is not present in any known language, they showed no effect of training (Dolscheid et al., submitted). These training studies demonstrate a causal role for language in strengthening the use of some nonlinguistic mappings more than others.

[1] The thickness-stimuli used in Experiment 2 could also be interpreted as a size manipulation. Indeed, even though movement was restricted to the horizontal plane which is characteristic for thickness, there is a concomitant difference in overall size. For the present purposes, nothing rests on being able to make the distinction between thickness and size, per se, since the reported inconsistency in the ability to make the cross-modal mapping to pitch applies equally to both spatial parameters.

While language may enforce particular pitch-space mappings, this proposal has to take into account that metaphors pose additional demands in language acquisition. Pitch metaphors are inherently polysemous; the acquisition of both spatial and sound meanings is likely more complex than when a single meaning has to be acquired (see e.g. Johnson, 1992). Consistent with this proposal French speaking children trained to describe sounds using either the single-meaning terms *aigu* and *grave* (a pair of antonyms used only to label high and low pitches) versus the polysemous words *haut* and *bas* (which are used to refer to pitch and space) were better able to label sound stimuli (Costa-Giomi & Descombes, 1996).

Aside from polysemy, another important attribute of metaphorical language lies in its directionality. Taking spatial metaphors of time as an example, people talk about time in terms of space far more often ("a long vacation"; "a short meeting") than they talk about space in terms of time (though it occasionally occurs: "I live two minutes from here") (Casasanto, 2008, 2010; Lakoff & Johnson, 1980). For pitch metaphors the same asymmetry seems to hold, which has also been found to be reflected in adults' nonlinguistic pitch representations (Casasanto, 2010). Note, however, that our results are agnostic of a space-pitch asymmetry in prelinguistic infants. While we have demonstrated that infants are able to detect space-pitch associations, our tasks do not speak to possible directionality. Future studies are necessary to determine whether language plays a role in introducing this asymmetry (e.g. see Merrit, Casasanto, & Brannon, 2010), or whether it is present independent of language (e.g. see Marks et al., 1987).

Effects of cross-modal associations on language

Since cross-modal associations are present before children acquire language, it is possible that the associations themselves shape metaphors in language. We find the height-pitch metaphor in languages such as Spanish, German and Polish (Rusconi et al., 2006), as well as non-Indo-European languages, like Japanese and Chinese. In all of these languages, "high" refers to high frequency sounds and "low" to low frequency sounds, but not the reverse. Likewise, psychophysical studies demonstrated that participants associate higher pitches with smaller objects, not with larger objects (e.g. Gallace & Spence, 2006). For the Kpelle and Jabo people in Liberia, this association is also encoded in language: "small" refers to high pitch and "big" refers to low pitch (see e.g., Eitan & Timmers, 2010). Prelinguistic associations, alongside correlations of properties in the real world, may thus serve as guiding principles that constrain the way pitch gets lexicalized, across languages.

Consequently, it might be harder to learn linguistic metaphors that are inconsistent with cross-modal mappings for which there is evidence in the natural world. The results of a training study support this suggestion. Dutch speakers trained to use reversed thickness-pitch mappings (thick=high, thin=low) were not able to master this association, even though they could learn the comparable congruent mapping (Dolscheid et al., submitted). It thus appears that language cannot easily retrain mappings that are supported by correlations present prelinguistically and/or supported by real world experience. Early sensitivity to certain mappings might therefore constrain the set of cross-modal associations that are likely to be observed in language and the mind.

Origins of cross-modal mappings?

Are cross-modal mappings innate? Based on the current evidence, we can only conclude that cross-modal associations between space and pitch are present from very early. By the age of 4 months, however, infants may well have encountered enough relevant co-occurrences in their interaction with the world to have learned these mappings. Thickness-pitch mappings seem especially prevalent: thicker strings produce lower notes, bigger bells have lower chimes, and people with bigger ('thicker') bodies tend to have lower voices. While infants may have internalized these regularities, the case for innate height-pitch mappings is not conclusive (see also Walker et al., 2010).

Conclusions

No matter whether cross-modal associations are inborn or learned, the finding that both height-pitch and thickness-pitch mappings can be observed in infants as young as 4 months of age constrains theorizing about the role that language plays in shaping nonlinguistic mental representations of pitch.

Our data show that space-pitch associations are present prior to language, suggesting that language is unlikely to create cross-modal mappings between space and pitch, even if language seems to play this role in other domains (Gentner, 2002).

It appears that both the height-pitch mapping found in languages like Dutch and the thickness-pitch mapping found in languages like Farsi are already present in prelinguistic infants' minds. This suggests that people who use different spatial metaphors for pitch in their native languages come to think about pitch differently not because language instills in them one cross-modal mapping instead of the other, but rather because language strengthens one pre-existing mapping at the expense the other, via some form of competitive associative learning (Casasanto, 2008, 2010; Dolscheid, et al., submitted). The precise learning mechanisms that give rise to cross-linguistic differences in pitch representation, and the underlying neural mechanisms, remain topics for future research.

Acknowledgments

We would like to thank Sho Tsuji, Manu Schuetze, Angela Khadar, Margret van Beuningen, Nienke Dijkstra, Laura Arendsen, Dirkje van der Aa & Webb Phillips.

References

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

Ben-Artzi, E., & Marks, L. E. (1995). Visual–auditory interaction in speeded classification: Role of stimulus difference. *Perception & Psychophysics, 57*, 1151-1162.

Bloom, P. (2000). How children learn the meanings of words. Cambridge, MA: MIT Press.

Bloom, P., & Keil, F. (2001). Thinking through language. *Mind and Language*, *16*, 351-367.

Braaten, R. (1993). *Synesthetic correspondence between visual location and auditory pitch in infants*. Paper presented at the 34th Annual Meeting of the Psychonomic Society.

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

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

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

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.

Dolscheid, S., Shayan, S., Majid, A., & Casasanto, D. (submitted). The Thickness of Musical Pitch**:** Psychophysical evidence for linguistic relativity.

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

Evans, K. K., & Treisman, A. (2010). Natural cross-modal mappings between visual and auditory features. *Journal of Vision*, *10(1)*, 1-12.

Gallace, A., & Spence, C. (2006). Multisensory synesthetic interactions in the speeded classification of visual size. *Perception & Psychophysics*, *68*, 1191-1203.

Gentner, D. (2002). Analogy in scientific discovery: The case of Johannes Kepler. In L. Magnani & N. J. Nersessian (Eds.), *Model-based reasoning: Science, technology, values* (pp.21-39). New York: Kluwer Academic/ Plenum Publisher.

Gentner, D. (2003). Why we're so smart. In D. Gentner and S. Goldin-Meadow (Eds.), *Language in mind: Advances in the study of language and thought* (pp.195-235). Cambridge, MA: MIT Press.

Gopnik, A., & Meltzoff, A. N. (1997). Words, Thoughts, and Theories. Cambridge, MA: MIT Press.

Hubbard, T. L. (1996). Synesthesia-like mappings of lightness, pitch, and melodic interval. *The American Journal of Psychology*, *109*, 219-238.

Jeschonek, S., Pauen, S., & Babocsai, L. (in preparation). Cross-modal mapping of visual and acoustic displays in infants: The effect of static and dynamic components.

Johnson, C. J. (1992). Cognitive components of naming in children: Effects of referential uncertainty and stimulus realism. *Journal of Experimental Child Psychology*, *53*, 45-71.

Lakoff, G., & Johnson, M. (1980). Metaphors we live by. Chicago and London: The University of Chicago Press.

Levinson, S. C., & Majid, A. (2007). The language of sound. In A. Majid (Ed.), Field Manual Volume 10 (pp. 29-31). Nijmegen: Max Planck Institute for Psycholinguistics.

Lewkowicz, D. J., & Turkewitz, G. (1980). Cross-modal equivalence in early infancy: Auditory–visual intensity matching. *Developmental Psychology*, *16*, 597-607.

Marks, L. E. (1987). On cross-modal similarity: Auditory–visual interactions in speeded discrimination. *Journal of Experimental Psychology and Human Perception Performance*, *13*, 384-394.

Marks, L. E. (1989). On cross-modal similarity: The perceptual structure of pitch, loudness, and brightness. *Journal of Experimental Psychology: Human Perception and Performance*, *15*, 586-602.

Marks, L. E., Hammeal, R. J., & Bornstein, M. H. (1987). Perceiving similarity and comprehending metaphor. *Monographs of the Society for Research in Child Development*, *52*(*1*), 1-102.

Martino, G., & Marks, L. E. (1999). Perceptual and linguistic interactions in speeded classification: Tests of the semantic coding hypothesis. *Perception*, *28*, 903-923.

Melara, R. D., & O'Brien, T. P. (1987). Interaction between synesthetically corresponding dimensions. *Journal of Experimental Psychology: General*, *116*, 323-336.

Melara, R. D. (1989). Dimensional interaction between color and pitch. *Journal of Experimental Psychology: Human Perception and Performance*, *15*, 69-79.

Merritt, D.J., Casasanto, D., & Brannon, E. M. (2010). Do Monkeys Think in Metaphors? Representations of Space and Time in Monkeys and Humans. *Cognition, 117*, 191-202.

Mondloch, C. J., & Maurer, D. (2004). Do small white balls squeak? Pitch–object correspondences in your children. *Cognitive, Affective & Behavioral Neuroscience*, *4*, 133-136.

Parise, C., & Spence, C. (2009). "When birds of a feather flock together": Synesthetic correspondences modulate audiovisual integration in non-synesthetes. *PLoS ONE*, *4*, e5664.

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

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

Shayan, S., Ozturk, O., Bowerman, M., & Majid, A. (submitted). Spatial metaphor in language can promote the development of cross-modal mappings in children.

Smith, L. B., & Sera, M. D. (1992). A developmental analysis of the polar structure of dimensions. *Cognitive Psychology*, *24*, 99-142.

Spence, C. (2011). Crossmodal correspondences: A tutorial review. *Attention, Perception & Psychophysics*, *73*, 971-995.

Roffler, S. K., & Butler, R. A. (1968). Factors that influence the localization of sound in the vertical plane. *The Journal of the Acoustical Society of America*, *43,* 1255-1259.

Wagner, K., & Dobkins, K. R. (2011). Synaesthetic Associations Decrease During Infancy. *Psychological Science*, *22,* 1067-1072.

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

Walker, P., Bremmner, 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.

Williams, J. M. (1976). Synaesthetic adjectives. A possible law of semantic change. *Language*, *52*, 461-478.