BRILL

Multisensory Research 29 (2016) 93–111

MULTISENSORY RESEARCH

brill.com/msr

Audio-Visual, Visuo-Tactile and Audio-Tactile Correspondences in Preschoolers

Elena Nava [1,3,*], **Massimo Grassi** [2] **and Chiara Turati** [1,3]

[1] University of Milan–Bicocca, Department of Psychology, Piazza dell'Ateneo Nuovo 1, 20126 Milan, Italy

[2] University of Padua, Via Venezia 8, 35131 Padua, Italy

[3] NeuroMI — Milan Centre for Neuroscience

Received 26 September 2014; accepted 15 March 2015

Abstract

Interest in crossmodal correspondences has recently seen a renaissance thanks to numerous studies in human adults. Yet, still very little is known about crossmodal correspondences in children, particularly in sensory pairings other than audition and vision. In the current study, we investigated whether 4–5-year-old children match auditory pitch to the spatial motion of visual objects (audio-visual condition). In addition, we investigated whether this correspondence extends to touch, i.e., whether children also match auditory pitch to the spatial motion of touch (audio-tactile condition) and the spatial motion of visual objects to touch (visuo-tactile condition). In two experiments, two different groups of children were asked to indicate which of two stimuli fitted best with a centrally located third stimulus (Experiment 1), or to report whether two presented stimuli fitted together well (Experiment 2). We found sensitivity to the congruency of all of the sensory pairings only in Experiment 2, suggesting that only under specific circumstances can these correspondences be observed. Our results suggest that pitch–height correspondences for audio-visual and audio-tactile combinations may still be weak in preschool children, and speculate that this could be due to immature linguistic and auditory cues that are still developing at age five.

Keywords

Crossmodal correspondences, audition, vision, touch, development, preschool children, learning

1. Introduction

Because the environment typically stimulates more than one sensory modality at the same time, the brain has to learn to give meaning to the incoming in-

* To whom correspondence should be addressed. E-mail: elena.nava@unimib.it

DOI:10.1163/22134808-00002493

formation by matching the 'right' stimuli, i.e., by determining which signals originate from the same or a different source and by detecting correspondences or mismatches between sensory attributes (Stein, 2012). Space and time represent the typical cues that the brain uses to integrate multisensory information (Stein *et al.*, 1988), but there is evidence that other factors contribute to crossmodal matching. For example, semantic congruence (Chen and Spence, 2010; Doehrmann and Naumer, 2008; Laurienti *et al.*, 2004) has been shown to modulate the processing of multisensory stimulation. In other words, if meaningful multisensory stimuli are used (e.g., an image of a dog coupled with the sound of a dog barking), the brain is facilitated in integrating the information (though see Koppen *et al.*, 2008).

Our brain also has the tendency to match certain features/dimensions of stimuli across different sensory modalities. These special mappings have been commonly labelled as 'crossmodal correspondences' (Parise and Spence, 2013; Spence, 2011), and broadly refer to redundant as well as non-redundant features. The term 'redundant' refers to one or more features of one stimulus that can be perceived through different sensory modalities. For example, the bounces produced from a ball striking a surface convey a rhythm that can be coded both from the auditory as well as the visual system. That is, the information conveyed — rhythm — is amodal because is not specific to a single sensory modality; thus, rhythm conveyed by seeing and hearing a ball bouncing is redundant. On the contrary, non-redundant crossmodal correspondences refer to associations of stimuli that do not share amodal information. For example, Melara (1989) showed that adults tend to associate luminance with pitch, i.e., 'white' with a high pitch, and 'black' with a low pitch.

Non-redundant crossmodal correspondences have been found in adults in various sensory pairings, ranging from the most investigated audio-visual correspondences (e.g., Evans and Treisman, 2010; Gallace and Spence, 2006; Parise and Spence, 2009) to audio-tactile correspondences (Occelli *et al.*, 2009), and from visuo-tactile correspondences (Ludwig and Simner, 2013; Martino and Marks, 2000; Slobodenyuk *et al.*, in press) to correspondences between vision and smell (Demattè *et al.*, 2006; Gilbert *et al.*, 1996) or sound and taste (Knöferle and Spence, 2012).

Research on non-redundant crossmodal correspondences has recently been extended to infants and children too to investigate the origins and development of such special mappings. However, these studies have reached contrasting results. For example, Walker and co-workers (2010) found that three-month-old infants looked longer at a ball rising and falling along a vertical trajectory when a concurrently presented frequency sweep was also rising and falling in pitch. Furthermore, the authors found that infants looked longer at an object with sharp edges when accompanied by a high pitched tone, rather than the same object accompanied by a low pitched tone, suggesting that visuo-spatial

pitch–height and visual-sharpness pitch–height correspondences may emerge very early in life. Dolscheid and co-workers (2014) replicated the height–pitch correspondence and also found, in another experiment, that four-month-old infants were sensitive to thickness–pitch correspondences. However, Lewkowicz and Minar (2014) recently failed to replicate the findings of Walker and co-workers (2010). Note that the authors conducted five experiments with three different age groups (three-, four-, and six-month-old infants), using the same stimuli as Walker and co-workers (2010), as well as different stimuli, but reached the conclusion that infants were not sensitive to crossmodal correspondences (though see the reply of Walker *et al.*, 2014).

Indeed, some types of crossmodal correspondences only emerge in later childhood. Haryu and Kajikawa (2012) found that ten-month-old infants were sensitive to brightness–pitch (i.e., infants associated a high frequency tone with an object of a brighter colour and a low frequency tone with an object of a darker colour), but not to size–pitch correspondences (i.e., infants did not always associate a larger object with a low frequency tone and a small object with a high frequency tone). Mondloch and Maurer (2004) presented children ranging from 30 to 35 months with a small white ball and a larger grey ball, bouncing in synchrony with each other along the vertical axis of the screen, together with a tone presented as the balls reversed their trajectory at the bottom of the screen. When the accompanying sound had a high frequency, most children pointed to the white and small ball. On the contrary, when the accompanying sound had a low frequency, most children pointed to the grey and larger ball. However, Smith and Sera (1992) showed that when asked to match the larger of two objects with the louder of two sounds (Exp. 3), two- to three-year-old children did not consistently match loud with big as adults do. Interestingly, Marks and co-workers (1987) found that a systematic matching between size and pitch did not appear before age 11 years.

It is worth noting that the contrasting results that emerged from studies of infants and young children are particularly relevant for one type of association, namely when pitch is associated with a spatial visual feature — be it size or height — which leads to the following questions: how is it that auditory pitch is so often associated with the spatial characteristics of the visual stimuli? How do children associate pitch with visual spatial features? Almost a century ago, Pratt (1930) observed that high-pitched tones are higher in space than low-pitched tones, and *vice versa*. Recently, Rusconi and co-workers (2006) found that auditory pitch is spatially represented. They showed that the spatial connotation of pitch interacted with the motor action. In one of their experiments, the participants were asked to respond quickly as to whether a pitch was high or low in comparison to a reference pitch. The participants proved to be faster whenever the response was coherent with the spatial position of the response key (i.e., the response was ‘higher’ and the response key was in

the upper part of the keyboard; the response was 'lower' and the response key was in the lower part of the keyboard). The authors thus suggested that pitch was spatially represented and that this spatial representation modulated motor reaction.

The spatial nature of auditory pitch has recently been further explored by Parise and co-workers (2014), who found that auditory scene statistics revealed a clear mapping between frequency and elevation (i.e., high-frequency sounds tend to originate from elevated sound sources). Most importantly, their findings suggested that the anatomy of the ear and how humans localise sounds in the environment evolved as a function of such statistics, suggesting that the pitch–elevation mapping may be an embodied system.

Indeed, as suggested by Parise and co-workers (2014), the origins of the association between pitch and space might be statistical and emerge because it is learned through the statistics of natural sounds. By the same token, natural languages may mirror this association (e.g., pitch is labelled as being 'low' or 'high' in different languages, including English, French, German, Italian, Spanish, Chinese and Polish). For instance, large objects tend to produce sounds rich in low frequencies, whereas small objects tend to produce sounds rich in high frequencies. Grassi (2005) and Grassi *et al*. (2013) examined whether adults could judge the size of a ball from the sound it produced when it is dropped upon a plate. The studies revealed that participants were able to estimate the size of the ball and that judgments were strongly correlated with (and influenced by) the distribution of the frequency content of the sounds produced by the ball. Melara and Marks (1990) tested the correspondences between linguistic and acoustical dimensions in a series of speeded classification tasks. In one of these experiments, the authors found that participants responded faster when the written word 'high' was accompanied by a high pitch, and when the written word 'low' was accompanied by a low pitch. These results suggest that human adults statistically learn certain audio-visual correspondences through the experience they have with objects and events in the environment (see Ernst, 2007 for further evidence on the possible role of statistical learning in the formation of crossmodal correspondences).

Given the particular spatial connotation of auditory pitch and also the controversial findings reported in infant and toddler studies (Lewkowicz and Minar, 2014; Mondloch and Maurer, 2004; Smith and Sera, 1992; Walker *et al.*, 2010), in the present study we investigated auditory pitch–visual height correspondences in preschool children aged between four and five years. There is very little literature on pre-literate children (though see Marks *et al.*, 1987; Smith and Sera, 1992), although it represents a very interesting age-group as it allows us to investigate whether these associations occur outside the influence of written language and musical training. Testing this particular age-group also allows us to use more adult-like methods (e.g., manual or verbal response, in

comparison to the preferential looking techniques used in infant studies), being sure at the same time that the children are capable of understanding the task. Moreover, this is the first study to explore whether crossmodal associations between auditory pitch and spatial height transfer to touch by testing children on three types of sensory pairings: audio-visual, audio-tactile and visuo-tactile. Tactile stimuli, as well as visual stimuli, carry spatial information. We wondered whether auditory pitch was associated with tactile stimuli to the same extent as visual stimuli.

To represent height within the visual sensory modality, we used the so-called 'Barber Pole illusion', a horizontal movement that is perceived as a vertical movement, therefore, a stimulus that moves illusorily along the vertical space. The auditory stimulus consisted of the so-called sonic Barber Pole, i.e., the Shepard–Risset glissando. The Shepard–Risset glissando is illusory and is perceived as continuously moving up (or down) in auditory pitch. These specific visual and audio stimuli were selected for the following reasons. The visual stimulus enabled us to represent one direction of motion at a time, moving continuously without interruption. The visual stimulus used in previous studies (i.e., a disc moving up and down along the vertical axis of the screen, e.g., Lewkowicz and Minar, 2014; Walker *et al.*, 2010) included both directions of motion as well as a change in direction of motion. By the same token, the auditory stimulus enabled us to represent one pitch direction at a time, continuously moving without interruption. The frequency sweep used in previous studies in fact (Lewkowicz and Minar, 2014; Walker *et al.*, 2010) included both pitch directions. In addition, in both studies (Lewkowicz and Minar, 2014; Walker *et al.*, 2010), the sweep was also modulated in loudness, while our stimuli were always kept constant in terms of loudness over time.

The tactile modality was represented by a manual touch delivered by an experimenter (blind to the purposes of the study) on the back of the participant by means of a paintbrush. The strokes were performed from the bottom to the top of the back (or *vice versa*). We chose manual strokes over tactile vibrations because we wanted to observe whether children were able to match pitch to another sense (other than vision) that moved in a spatial dimension. For this reason, the visuo-tactile stimulus somehow served as a control, as both visual and tactile stimuli move along the same spatial dimension (i.e., height), thus conveying redundant information about the direction of motion.

The current study was constituted by two experiments. In the first experiment children were asked to decide which of two lateralised stimuli (e.g., two visual stimuli moving in opposite directions) would fit best with a third stimulus of a different sensory modality (e.g., an auditory stimulus moving congruently only with one of the two visual stimuli). In the second experiment, children were asked to judge whether two different modality stimuli (e.g., a visual barber pole and a concurrent auditory barber pole) were congruent or

not. We hypothesised that if the children were sensitive to all sensory combinations, they would match the stimuli according to direction of movement and thus prefer congruent over incongruent pairings. In both experiments, the results of the children were compared to the results of two groups of adult participants.

2. Experiment 1

2.1. Method

2.1.1. Participants

Fifteen children aged 4–5 years (five females, mean age = 5.0 years, age range: 4.2–5.7 years) were recruited from an Italian nursery school (Scuola dell'Infanzia 'Don Bosco', Cesate, Italy). All of them had normal or corrected-to-normal vision and hearing.

Twenty naïve adults also took part in the experiment (seven females, mean age = 30.7 years, range 22–40 years of age), and all received a certificate for their participation. All had normal or corrected-to-normal vision and hearing.

None of the participants presented cognitive/neurological impairment. Furthermore, none of the participants were informed about the hypothesis of the study prior to testing or had previous experience with musical training.

The adult participants signed an informed consent before the beginning of the experiment. For the children, informed consent was signed by both parents before testing. The study was approved by the Ethical Committee of the University of Milan–Bicocca (Italy).

2.1.2. Apparatus and Materials

The auditory and visual stimuli were presented on a 15.6-inch Dell Inspiron laptop. The visual stimuli consisted of sinusoidal gratings alternately white and red moving behind an elongated rectangular aperture, giving the impression of a movement along the major axis of the aperture (the so-called 'barber-pole illusion'). The stimuli were presented on a black background. The total field size of the aperture display was 6.7 degrees of visual angle, the spatial frequency of the gratings was 0.8 cpd, and they moved at ca. 0.75°/s. The gratings inside the aperture were oriented at 45° and moved horizontally to the right or to the left giving the impression of vertical motion (i.e., moving down or up, respectively, see Fig. 1a and online Supplementary movie [mov] file with visual stimuli).

The auditory stimuli consisted of several tones gliding continuously in frequency (i.e., the so-called 'Shepard–Risset glissando', a variation of the 'Shepard scale', Shepard, 1964). The stimulus consisted of a superposition of nine harmonic tones each gliding log-linearly in frequency. The amplitude of each sine wave was modulated with a bell-shape Gaussian envelope thus

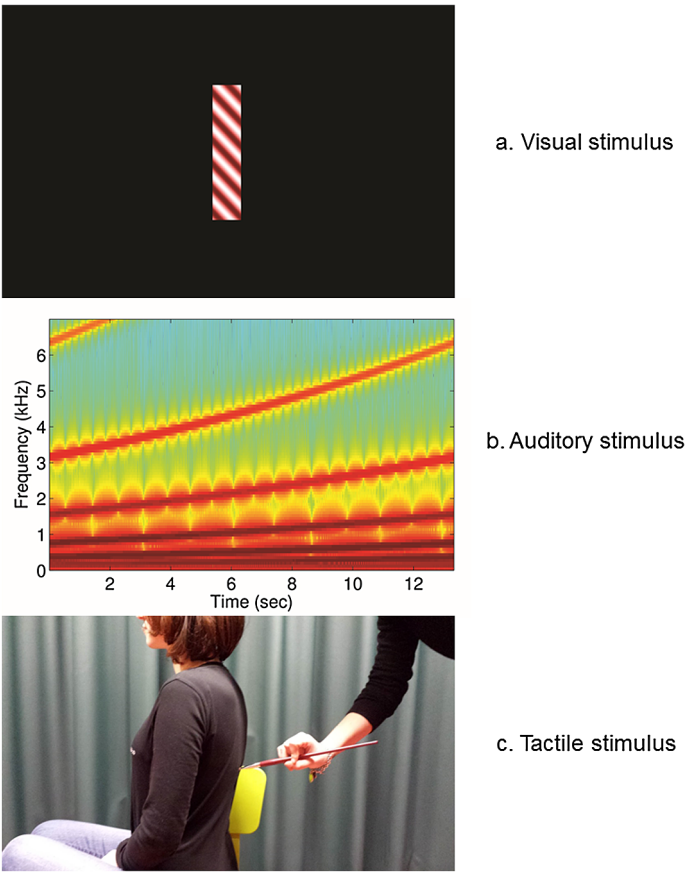

Figure 1. One frame of the visual stimulus, the visual barber-pole (a); spectrogram of the Shepard–Risset glissando (b) and a photo depicting how the tactile stimulus was delivered to the participant (c). This figure is published in colour in the online version.

creating the auditory illusion of a complex tone that continually rose (or fell) in pitch (also called the 'sonic barber-pole'). The frequencies of the stimulus ranged between 27.5 and 8000 Hz and the stimulus was presented at the constant overall level of 65 dB SPL (measured from the head of the participant; see Fig. 1b and online Supplementary sound [wav] recording). The visual and auditory stimuli were programmed in Matlab 2013a with custom Matlab toolboxes (Grassi and Soranzo, 2009; Kleiner *et al.*, 2007).

The tactile stimuli consisted of a paintbrush that was used to stroke the back of the infant. The movements were performed from the lowest part of the back to the neck or from the neck to the lowest part of the back (i.e., the motion produced by the tactile stimulus was discrete, in that the paintbrush was always moved either upwards or downwards by the experimenter within

a single trial to provide the sense of tactile motion in a single direction; see Fig. 1c, and online Supplementary video [mp4] file).

The crossmodal correspondences presented were visuo-tactile, audio-tactile and audio-visual. Each condition (i.e., visuo-tactile, audio-tactile and audio-visual) consisted of four trials, two of which were congruent, whereas the remaining two were incongruent. 'Congruent' meant that the direction of motion was similar in the two sensory modalities, e.g., in the audio-visual condition, the red and white gratings moved upwards and the pitch of the Shepard–Risset glissando rose; in the visuo-tactile condition, the red and white gratings moving upwards and the stroking was performed from the lower part of the back to the neck; in the audio-tactile condition, the Shepard–Risset glissando rose in pitch and the stroking was performed from the lower part of the back to the neck. 'Incongruent' meant that the direction of motion was opposite in the two sensory modalities, e.g., in the audio-visual condition, the red and white gratings moved upwards and the Shepard–Risset glissando fell in pitch; in the visuo-tactile condition, the red and white gratings moved downwards and the stroking was performed from the lower part of the back to the neck; in the audio-tactile condition, the Shepard–Risset glissando fell in pitch and the stroking was performed from the lower part of the back to the neck.

In the audio-visual condition, two visual stimuli were presented laterally on the right and left side of the monitor and only one Shepard–Risset glissando was concurrently presented. Note that auditory and visual stimuli, contrary to the tactile stimulus, were continuous and were presented from approximately the same spatial position. In the visuo-tactile condition, two visual stimuli were laterally presented on the right and left side of the monitor and only one stroke was concurrently performed. In the audio-tactile condition, two strokes were performed and only one Shepard–Risset glissando was concurrently presented.

2.1.3. Procedure

The adult controls were tested in a quiet room at the University of Milan–Bicocca.

All of the children were tested in a quiet room provided by the nursery school. The children were asked to sit in front of the monitor keeping a distance of approximately 60 cm. To familiarise the child with the experiment, the experimenter showed six slides to each child in which two images (e.g., the image of a dog, a cat, a train, etc.) were presented on the right and left side of the screen, one of which was congruently/semantically coupled with a corresponding stimulus (e.g., the barking of a dog coupled with the image of a dog). After each slide, the experimenter asked the child to point to the image s/he felt fitted best with the concurrent stimulus. All of the children were able to point to the congruent stimuli. Successively, they were told that they had to

perform a similar task with rectangles containing white and red stripes, continuous tones and strokes performed on their back. All of the children met the criteria for inclusion in the experiment (i.e., correct matching of audio-visual stimuli on a minimum of three slides).

The design was semi-randomised, i.e., half the children started with a congruent trial and the other half with an incongruent trial. For each condition, the question was always the same, i.e., to point to the stimulus that fitted best with the concurrent stimulus. More precisely, in the audio-visual condition, the children were asked the following question: 'Which of the two rectangles do you think fits best with the sound you are hearing now? The one on the right side or the one on the left side?'. In the visuo-tactile condition, children were asked: 'Which of the two rectangles do you think fits best with the stroking you are feeling on your back? The one on the right side or the one on the left side?'. In the audio-tactile condition, children were asked: 'Which of the two strokes do you think fits best with the sound you are hearing now? The one on the right side or the one on the left side?'. For the audio-visual and visuo-tactile conditions, the mentioning of the side was accompanied by a gesture pointing to either the right or left stimulus presented on the laptop. In the audio-tactile condition, the mentioning of the side was accompanied by a gesture pointing to either the right or left side of the back. The question was always asked during the presentation of the stimuli (approximately after 5 s from stimuli onset), which had a maximum duration of 60 s. For all of the children, this time was sufficient to provide an answer.

Note that there were always two experimenters performing the test. One of these was responsible for administering and recording the trials for each sensory pairing. The other experimenter was responsible for the tactile stimulation and was blind to the purpose of the study. Furthermore, we did not provide any training on the type of stimuli shown nor did we show the stimuli unimodally prior to testing because we wanted to see whether the children could automatically and spontaneously match the two stimuli. For the same reason, we kept the number of trials to a minimum (i.e., four trials, to avoid learning effects). No feedback was provided as to whether the answer was correct or incorrect (in fact, there is no correct answer in this task).

At the end of the experiment, each child received a sticker as a gift for his/her participation.

The adults were tested with the same method, with the exception of the slides (they were not presented) and the question, in that we asked adults to point to the concurrent congruent stimulus.

2.2. Results

We calculated the proportion of congruent responses for each child, with 'congruent' indicating responses given in the expected direction (e.g., Ris-

Table 1.
Proportion of congruent responses of each child in Experiment 1, separately for the audio-visual (AV), visuo-tactile (VT) and audio-tactile (AT) conditions

Participant	Condition		
	Audio-visual	Visuo-tactile	Audio-tactile
1	1.00	0.75	0.00
2	0.00	1.00	0.00
3	0.50	0.50	0.75
4	0.50	1.00	0.50
5	1.00	1.00	1.00
6	0.75	1.00	1.00
7	0.25	0.00	0.75
8	1.00	1.00	0.50
9	0.25	0.25	0.25
10	0.00	0.25	0.00
11	0.75	0.75	1.00
12	0.50	0.50	0.50
13	0.50	1.00	1.00
14	1.00	1.00	1.00
15	0.00	1.00	1.00
Mean	0.53	0.73	0.62
St. Dev.	0.38	0.35	0.40

set glissando going up in pitch and Barber pole moving up in space) for each sensory combination (audio-visual, visuo-tactile and audio-tactile).

The adults obtained 100% of congruent responses in all conditions.

The children reported 53% of congruent responses in the audio-visual condition, 73% in the visuo-tactile condition and 62% in the audio-tactile condition (see Table 1 for details about performance of each child).

Because the data were not normally distributed, we analysed the data with non-parametric tests.

To observe whether the children were sensitive to crossmodal correspondences, we tested the proportion of congruent responses, separately for the audio-visual, visuo-tactile and audio-tactile condition against 50% (chance response) using Wilcoxon Signed Rank Tests. The children's performance was above chance in the visuo-tactile condition ($z = 2.3$, $p = 0.02$) but not in the audio-visual ($p = 0.7$) and audio-tactile ($p = 0.3$) conditions (see Fig. 2).

We also compared the visuo-tactile condition between children and adults (with Mann–Whitney tests) and found that the children's proportion of congruent responses was lower than that of the adults ($z = -3.34$, $p = 0.001$).

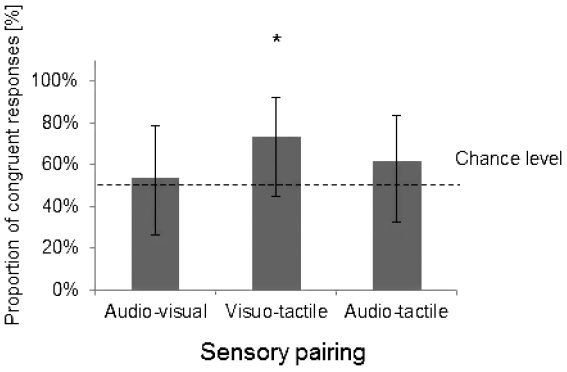

Figure 2. Proportion of congruent responses for the three sensory pairings (audio-visual, visuo-tactile and audio-tactile) in Experiment 1. Asterisks indicate performance above 50%. Error bars represent 95% binomial proportion confidence interval.

Furthermore, we also controlled whether the children's performance could be explained by a preference for side (i.e., right or left stimulus), but comparison of congruent responses for left *vs.* right side (with a related-samples Wilcoxon Signed Rank Test) did.

2.3. Interim Discussion

In Experiment 1, we found that the children were able to detect the crossmodal correspondences in the visuo-tactile, but not in the audio-visual and in the audio-tactile conditions.

One possible explanation of children's inability to match audio-visual and audio-tactile stimuli may reside in the paradigm used, in which three different stimuli were shown simultaneously (i.e., one auditory stimulus and two visual stimuli in the audio-visual condition; one auditory stimulus and two tactile stimuli in the audio-tactile condition). Children may have had difficulties to isolate and process a pair of stimuli impinging on two different modalities, while inhibiting interference from a third stimulus. In other words, our design may have reflected children's limitations in terms of memory, inhibition and cognitive flexibility (see Davidson *et al.*, 2006), rather than insensitivity to crossmodal correspondences. In the visuo-tactile condition, because the visual and tactile stimuli were moving along the same spatial dimension, children may have experienced as redundant the same direction of motion, thus being facilitated in the task.

Following this line of reasoning, in Experiment 2, we tested a new group of children of the same age on a modified version of the paradigm, in which the children were stimulated with only two stimuli at a time (e.g., a tactile stimulus coupled with a visual stimulus).

3. Experiment 2

3.1. Method

3.1.1. Participants

Fourteen children aged 4–5 years (sevenfemales, mean age = 5.1 years, age range: 4.1–5.9 years) were recruited from the 'Scuola dell'Infanzia Don Bosco' (Cesate, Italy) to take part in this study. All of the children had normal or corrected-to-normal vision and hearing.

Ten students were recruited from the University of Milan–Bicocca (five females, mean age = 23.2 years, age range: 20–26 years) and received a course credit for their participation. All of them had normal or corrected-to-normal vision and hearing. None of the participants had cognitive or neurological impairment. All of them were naïve as to the purpose of the experiment prior to the testing and had no previous experience with musical training. The adult participants signed an informed consent prior to the beginning of the experiment. For the children, informed consent was signed by both parents before testing. The study was approved by the Ethical Committee of the University of Milan–Bicocca (Italy).

3.1.2. Apparatus and Materials

The stimuli used in Experiment 2 were the same as in Experiment 1. In this version of the experiment, we manipulated the type of stimuli presentation. In other words, contrary to Experiment 1, the participants viewed one pair of stimuli at a time. For example, in the audio-visual condition, all of the participants only viewed a single visual stimulus accompanied by a Shepard–Risset glissando and were asked to provide a 'yes/no' response following the question as to whether they considered the two stimuli to fit well together. Similarly to Experiment 1, the total number of trials was kept to four, including two congruent and two incongruent trials, and the order of presentation was counterbalanced among participants.

3.1.3. Procedure

The adult controls were tested in a quiet room of the University of Milan–Bicocca.

All of the children were tested in a quiet room provided by the nursery school. The children were asked to sit in front of the monitor at a distance of approximately 50/60 cm. To familiarise the child with the experiment, the experimenter showed up to six slides to each child in which images (e.g., the image of a dog, a cat, a train, etc.) were coupled with the (semantically) corresponding sound (e.g., the barking of a dog coupled with the image of a dog) or a non-corresponding sound (e.g., the barking of the dog coupled with the image of a train). After each slide, the question posed to the child was always the same: 'Do you think the image fits well with the sound?'. All of the chil-

dren were able to congruently match the image with the sound by providing a 'yes/no' response and were then told that they would now have to perform a similar task with rectangles containing white and red stripes, continuous tones and strokes performed on their back. The duration of the trials was the same as in Experiment 1 and all children were able to provide an answer within the time given. As in Experiment 1, two experimenters performed the test, with one performing the tactile stimulation, blind to the purposes of the study. The latter was also different from the experimenter performing the tactile stimulation in Experiment 1. At the end of the experiment, each child received a sticker as a gift for his/her participation.

The adults were tested with the same method, with the exception of the slides, which were not shown.

3.2. Results

The analyses for this experiment mimicked the analyses performed in Experiment 1. For each participant we calculated the proportion of congruent responses, obtained by assigning a point to each trial in which the child matched the stimuli in the expected direction (i.e., saying 'yes' when the stimuli were moving in the same direction; saying 'no' when the stimuli were moving in opposite directions).

In all conditions, all adults had 100% of congruent responses.

The children reported 84% of congruent responses in the audio-visual condition, 98% in the visuo-tactile condition and 89% in the audio-tactile condition (see Table 2 for single performance). This time, children proved sensitive to crossmodal correspondences in all sensory combinations (all $p < 0.02$, on Wilcoxon Signed Rank Test, see Fig. 3). Furthermore, the congruent responses of children did not differ from those of adults in the visuo-tactile and audio-tactile conditions (both $p > 0.1$), but did in the audio-visual condition ($z = 2.6$, $p = 0.01$).

Finally, we compared the three different sensory combinations in children and found differences between them (on Friedman's test, $x^2(2) = 6.5$, $p = 0.04$), caused by greater sensitivity to visuo-tactile correspondences than audio-visual correspondences ($z = 2.3$, $p = 0.02$).

4. General Discussion

The present study investigated crossmodal correspondences in four- to five-year-old children and two groups of adults for three different sensory pairings (audio-visual, visuo-tactile and audio-tactile). In particular, we investigated whether children were sensitive to the crossmodal correspondence that associates spatial height (either visual or tactile) with auditory pitch height. The literature shows mixed results for this correspondence in the developmental

Table 2.
Proportion of congruent responses of each child in Experiment 2, separately for the audio-visual (AV), visuo-tactile (VT) and audio-tactile (AT) conditions

Participant	Condition		
	Audio-visual	Visuo-tactile	Audio-tactile
1	1.00	1.00	0.75
2	1.00	1.00	1.00
3	0.75	1.00	1.00
4	1.00	1.00	1.00
5	1.00	1.00	0.75
6	1.00	1.00	1.00
7	0.50	1.00	0.50
8	0.50	1.00	0.75
9	0.75	1.00	1.00
10	1.00	1.00	1.00
11	0.75	1.00	0.75
12	1.00	1.00	1.00
13	0.75	0.75	1.00
14	0.75	1.00	1.00
Mean	0.84	0.98	0.89
St. Dev.	0.19	0.07	0.16

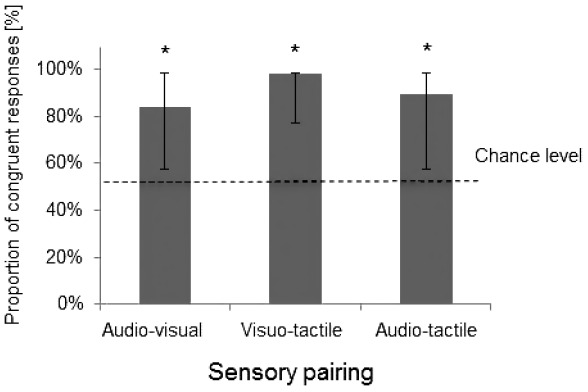

Figure 3. Proportion of congruent responses for the three sensory pairings (audio-visual, visuo-tactile and audio-tactile) in Experiment 2. Asterisks indicate performance above 50%. Error bars represent 95% binomial proportion confidence interval.

population (Dolscheid *et al.*, 2014; Lewkowicz and Minar, 2014; Walker *et al.*, 2010). In addition, these mixed results were gathered with auditory and visual stimuli that might have caused difficulty in the interpretation of the results (i.e., change in direction of motion for the visual stimulus, change in pitch and concurrent change in loudness for the auditory stimulus). For this reason, in our study we selected audio and visual stimuli that produced the illusion of continuously moving in only one direction (either in spatial height or in auditory pitch) and were not modulated in loudness over time.

In our study, we found that when asked to decide which of two stimuli congruently fitted with another stimulus presented in a different sensory modality (Experiment 1), the children consistently chose the congruent matching only in the visuo-tactile condition. However, when presented with only two stimuli at a time (Experiment 2), the children proved sensitive to all sensory combinations. That is, in Experiment 2 the children showed the ability to extract a rule for matching congruent stimuli (i.e., direction of motion).

Overall, our study shows two main findings. First, the children were sensitive to auditory pitch and visual and tactile height, but only under specific experimental situations. Second, the visuo-tactile combination appeared to be the 'easiest' for the children and was not influenced by the testing procedure (i.e., presenting two or three stimuli at a time).

As a note of caution, we should mention that the experimenters responsible for tactile stimulation could see the visual stimuli, and hear the auditory ones. As a consequence, it could be claimed that the experimenters were affected by the illusions, which in turn might have influenced the way in which the tactile stimulation was provided (e.g., by involuntarily putting more pressure or speeding up the tactile stimulation in the direction matching the visual or the auditory stimulus, thus making a particular tactile stimulus more salient over another one). Future studies should introduce a mechanical tactile stimulation system, in order to avoid any human bias. Nonetheless, we believe that this explanation is not tenable because the experimenters were naïve to the purpose of the study. A more plausible explanation as to why the children were particularly sensitive to the visuo-tactile combination probably resides in the stimuli used to investigate this correspondence. Indeed, the visual stimulus and the tactile strokes were both moving along the same spatial dimension, and the children probably perceived the congruent direction of motion as redundant. In other words, the children extracted the redundant amodal characteristics of the two stimuli (i.e., motion) and then matched the congruent direction of motion. Indeed, there is a lot of evidence supporting the notion that amodal information is the cornerstone of early perceptual development (e.g., Bahrick *et al.*, 2004; Lewkowicz, 2000), and that even infants can extract such information from multimodal stimulation.

By the same token, it could be argued that the children had more difficulties matching the pitch with the visual and tactile height because the pitch of the auditory stimulus varied within a tonal space (i.e., in frequency). However, they did match congruent pitch and height in Experiment 2, which leads to the question: how do children match these characteristics?

According to the interpretation of previous studies (Maurer *et al.*, 2006; Mondloch and Maurer, 2004; Walker *et al.*, 2014), these mappings could be a remnant of cortical interconnections already present at birth that did not undergo pruning. However, this hypothesis cannot fully explain the selective preference for congruent over incongruent stimuli.

There are other explanations that should be taken into consideration. The first refers to the possibility that children may have attributed a linguistic label to high and low frequencies. The children may have attributed the labels 'up' and 'down' to the pitches producing an illusory motion either upwards or downwards, respectively. Although not exposed to any musical training, we cannot exclude that they may have heard these words associated with pitch characteristics (precisely because denoting pitches as being 'low' or 'high' is common to many languages and in everyday speech).

The second explanation refers to the possibility that children are born with some predispositions favouring the learning of pitch–height associations. Indeed, Parise and co-workers (2014) found that there is a 'natural' mapping between frequency and elevation, provided by the fact that high-frequency sounds tend to originate from elevated sources. Most importantly, the authors also suggested that the shape of the human ear might has evolved to mirror the acoustic properties of the natural environment. Furthermore, Rusconi and co-workers (2006) found that there is a mental spatial representation of pitch, by which we naturally tend to associate high pitches with high elevation and low pitches with low elevation.

If pitch possesses an intrinsic spatial characteristic, it means that when matched with congruent visual spatial stimuli (e.g., high pitch with a high visual stimulus), they are processed as redundant in the brain. Translated to our experiment, this would mean that motion was coded as amodal information by auditory, visual and tactile stimuli, thus facilitating perception of congruency when the stimuli were perceived as moving in the same direction.

In line with the account provided by Parise and co-workers (2014), it could be speculated that the head-related transfer function (HRTF, which describes how a sound is filtered by the diffraction and reflection properties of the head, pinna and torso) is still refining in preschool children. In other words, because the head, pinna and torso are still developing at age five years (see Fels *et al.*, 2004 for evidence of differences between children and adult HRTF), it could be speculated that children's ears and localisation behaviour are still

fine-tuning the statistics of auditory scenes and that their ability to match pitch with height is concurrently developing too.

This hypothesis might contribute to explain why certain experimental procedures fail to grasp children's ability to detect crossmodal correspondences that involve auditory stimuli at age 4–5 years. Indeed, weakness in perceiving pitch–height correspondences would also explain the contrasting results found in infants (Dolscheid *et al.*, 2014; Lewkowicz and Minar, 2014; Walker *et al.*, 2010, 2014). In other words, it may be that sensitivity to this correspondence is not present at birth, as previously suggested (Walker *et al.*, 2010), but that it is learned by auditory experience. In turn, the anatomy of the ear predisposes young children to make/prefer some associations over others. Finally, language may additionally fine-tune height–pitch associations.

In conclusion, our study showed that preschool children were sensitive to audio-visual, visuo-tactile and audio-tactile correspondences, but only under specific experimental conditions, suggesting that an adult-like sensitivity to these correspondences is still developing by age five years.

Acknowledgements

We would like to thank all of the children who participated in the study. A special thanks goes to the educators of the 'Scuola dell'Infanzia Don Bosco' (Cesate, Italy) that made this study possible. This study was supported by a European Research Council Starting Grant to C.T. on a project entitled 'The origins and development of the human mirror neuron system' — ODMIR No. 241176.

References

Bahrick, L. E., Lickliter, R. and Flom, R. (2004). Intersensory redundancy guides the development of selective attention, perception, and cognition in infancy, *Curr. Dir. Psychol. Sci.* **13**, 99–102.

Chen, Y.-C. and Spence, C. (2010). When hearing the bark helps to identify the dog: semantically-congruent sounds modulate the identification of masked pictures, *Cognition* **114**, 389–404.

Davidson, M. C., Amso, D., Anderson, L. C. and Diamond, A. (2006). Development of cognitive control and executive functions from 4 to 13 years: evidence from manipulations of memory, inhibition, and task switching, *Neuropsychologia* **44**, 2037–2078.

Dematte, M. L., Sanabria, D. and Spence, C. (2006). Cross-modal associations between odors and colors, *Chem. Senses* **31**, 531–538.

Doehrmann, O. and Naumer, M. J. (2008). Semantics and the multisensory brain: how meaning modulates processes of audio-visual integration, *Brain Res.* **1242**, 136–150.

Dolscheid, S., Hunnius, S., Casasanto, D. and Majid, A. (2014). Prelinguistic infants are sensitive to space-pitch associations found across cultures, *Psychol. Sci.* **25**, 1256–1261.

Ernst, M. O. (2007). Learning to integrate arbitrary signals from vision and touch, *J. Vis.* **7**, 1–14.

Evans, K. K. and Treisman, A. (2010). Natural cross-modal mappings between visual and auditory features, *J. Vis.* **10**, 1–12.

Fels, J., Buthmann, P. and Vorländer, M. (2004). Head-related transfer functions of children, *Acta Acust. United Acust.* **90**, 918–927.

Gallace, A. and Spence, C. (2006). Multisensory synesthetic interactions in the speeded classification of visual size, *Percept. Psychophys.* **68**, 1191–1203.

Gilbert, A. N., Martin, R. and Kemp, S. E. (1996). Cross-modal correspondence between vision and olfaction: the color of smells, *Am. J. Psychol.* **109**, 335–351.

Grassi, M. (2005). Do we hear size or sound? Balls dropped on plates, *Percept. Psychophys.* **67**, 274–284.

Grassi, M. and Soranzo, A. (2009). MLP: a MATLAB toolbox for rapid and reliable auditory threshold estimations, *Behav. Res. Methods* **41**, 20–28.

Grassi, M., Pastore, M. and Lemaitre, G. (2013). Looking at the world with your ears: how do we get the size of an object from its sound? *Acta Psychol.* **143**, 96–104.

Haryu, E. and Kajikawa, S. (2012). Are higher-frequency sounds brighter in color and smaller in size? Auditory–visual correspondences in 10-month-old infants, *Infant Behav. Dev.* **35**, 727–732.

Kleiner, M., Brainard, D., Pelli, D., Ingling, A., Murray, R. and Broussard, C. (2007). What's new in Psychtoolbox-3, *Perception* **36**, ECVP Abstract Suppl. 1.

Knöferle, K. and Spence, C. (2012). Crossmodal correspondences between sounds and tastes, *Psychon. B. Rev.* **19**, 992–1006.

Koppen, C., Alsius, A. and Spence, C. (2008). Semantic congruency and the Colavita visual dominance effect, *Exp. Brain Res.* **184**, 533–546.

Laurienti, P. J., Kraft, R. A., Maldjian, J. A., Burdette, J. H. and Wallace, M. T. (2004). Semantic congruence is a critical factor in multisensory behavioral performance, *Exp. Brain Res.* **158**, 405–414.

Lewkowicz, D. J. (2000). The development of intersensory temporal perception: an epigenetic systems/limitations view, *Psychol. Bull.* **126**, 281–308.

Lewkowicz, D. J. and Minar, N. J. (2014). Infants are not sensitive to synesthetic cross-modality correspondences. A comment to Walker et al. (2010), *Psychol. Sci.* **25**, 832–834.

Ludwig, V. U. and Simner, J. (2013). What colour does that feel? Tactile–visual mapping and the development of cross-modality, *Cortex* **49**, 1089–1099.

Marks, L. E., Hammeal, R. and Bornstein, M. (1987). Perceiving similarity and comprehending metaphor, *Monogr. Soc. Res. Child Dev.* **52**, 1–102.

Martino, G. and Marks, L. E. (2000). Cross-modal interaction between vision and touch: the role of synesthetic correspondence, *Perception* **29**, 745–754.

Maurer, D., Pathman, T. and Mondloch, C. J. (2006). The shape of boubas: sound–shape correspondences in toddlers and adults, *Dev. Sci.* **9**, 316–322.

Melara, R. D. (1989). Dimensional interaction between color and pitch, *J. Exp. Psychol., Hum. Percept. Perform.* **15**, 69–79.

Melara, R. D. and Marks, L. E. (1990). Processes underlying dimensional interactions: correspondences between linguistic and nonlinguistic dimensions, *Mem. Cognit.* **18**, 477–495.

Mondloch, C. J. and Maurer, D. (2004). Do small white balls squeak? Pitch-object correspondences in young children, *Cogn. Affect. Behav. Neurosci.* **4**, 133–136.

Occelli, V., Spence, C. and Zampini, M. (2009). Compatibility effects between sound frequency and tactile elevation, *Neuroreport* **20**, 793–797.

Parise, C. and Spence, C. (2009). 'When birds of a feather flock together': synesthetic correspondences modulate audiovisual integration in non-synesthetes, *PLoS One* **4**, e5664. DOI:10.1371/journal.pone.0005664.

Parise, C. V. and Spence, C. (2013). Audiovisual crossmodal correspondences, in: *The Oxford Handbook of Synesthesia*, J. Simner and E. M. Hubbard (Eds), pp. 790–815. Oxford University Press, Oxford, UK.

Parise, C., Knorre, K. and Ernst, M. O. (2014). Natural auditory scene statistics shapes human spatial hearing, *Proc. Natl Acad. Sci. USA* **111**, 6104–6108.

Pratt, C. C. (1930). The spatial character of high and low tones, *J. Exp. Psychol.* **13**, 278.

Rusconi, E., Kwan, B., Giordano, B. L., Umilta, C. and Butterworth, B. (2006). Spatial representation of pitch height: the SMARC effect, *Cognition* **99**, 113–129.

Shepard, R. N. (1964). Circularity in judgments of relative pitch, *J. Acoust. Soc. Am.* **36**, 2346–2353.

Slobodenyuk, N., Jraissati, Y., Kanso, A., Ghanem, L. and Elhajj, I. (in press). Cross-modal associations between color and haptics, *Atten. Percept. Psychophys.*, DOI: 10.3758/s13414-015-0837-1.

Smith, L. B. and Sera, M. D. (1992). A developmental analysis of the polar structure of dimensions, *Cogni. Psychol.* **24**, 99–142.

Spence, C. (2011). Crossmodal correspondences: a tutorial review, *Atten. Percept. Psychophys.* **73**, 971–995.

Stein, B. E. (Ed.) (2012). *The New Handbook of Multisensory Processing*. MIT Press, Cambridge, MA, USA.

Stein, B. E., Huneycutt, S. W. and Meredith, A. (1988). Neurons and behavior: the same rules of multisensory integration apply, *Brain Res.* **448**, 355–358.

Walker, P., Bremner, J. G., Mason, U., Spring, J., Mattock, K., Slater, A. and Johnson, S. P. (2010). Preverbal infants' sensitivity to synaesthetic cross-modality correspondences, *Psychol. Sci.* **21**, 21–25.

Walker, P., Bremner, J. G., Mason, U., Spring, J., Mattock, K., Slater, A. and Johnson, S. P. (2014). Preverbal infants are sensitive to crossmodal correspondences. Much ado about the null results of Lewkowicz and Minar (2014), *Psychol. Sci.* **25**, 835–836.