



Citation for published version:

Petrini, K, Jones, PR, Smith, L & Nardini, M 2015, 'Hearing where the eyes see: children use an irrelevant visual cue when localizing sounds', *Child Development*, vol. 86, no. 5, pp. 1449-1457.
<https://doi.org/10.1111/cdev.12397>

DOI:

[10.1111/cdev.12397](https://doi.org/10.1111/cdev.12397)

Publication date:

2015

Document Version

Peer reviewed version

[Link to publication](#)

This is the pre-peer reviewed version of the following article: Petrini, K., Jones, P. R., Smith, L., & Nardini, M. (2015). Hearing where the eyes see: children use an irrelevant visual cue when localizing sounds. *Child Development*, 86(5), 1449-1457. 10.1111/cdev.12397, which has been published in final form at 10.1111/cdev.12397. This article may be used for non-commercial purposes in accordance with Wiley Terms and Conditions for Self-Archiving.

University of Bath

General rights

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

Take down policy

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

Hearing where the eyes see: children use an irrelevant visual cue when localizing
sounds

Karin Petrini^{1,2}, Peter R Jones², Louise Smith³ and Marko Nardini^{4,2}

¹Department of Psychology, University of Bath, UK

²Institute of Ophthalmology, University College London, UK

³Department of Experimental Psychology, University of Oxford, UK

⁴Department of Psychology, Durham University

Format: Research Article

Corresponding Author:

Dr Karin Petrini

Lecturer

Department of Psychology

University of Bath

Claverton Down

Bath

BA2 7AY

United Kingdom

Email: K.Petrini@bath.ac.uk

Running Head: Children over-weight irrelevant multisensory cues

Hearing where the eyes see: children use an irrelevant visual cue when localizing
sounds

Abstract

To reduce sensory uncertainty, humans combine cues from multiple senses. However, in everyday life, many co-occurring cues are irrelevant to the task at hand. How do humans know which cues to ignore? And does this ability change with development? Here we show that the ability to ignore cross-modal irrelevant information develops late in childhood. Participants performed a sound discrimination task, with or without an irrelevant visual flash, presented synchronously in front of them. Adults ignored the irrelevant visual information, while 7- to 10-year-olds responses were biased towards the flash location. Our findings show that acquiring mature cue combination mechanisms is a multifaceted process that includes learning to ignore irrelevant cues, as well as to optimally combine relevant cues.

We live in a multisensory world and our perception of it can be greatly improved by integrating multiple sensory cues to reduce sensory uncertainty (Ernst & Banks, 2002; Landy, Maloney, Johnston, & Young, 1995). For example, when crossing a road we can use both the sight and sound of an approaching car to best estimate its position. However, our perception can be deceived if we combine sensory cues that we should not. Magicians have exploited this knowledge for centuries to perform ventriloquism, a stagecraft act in which the voice coming from a person (the ventriloquist) is actually perceived as coming from another source (a puppet moving its mouth together with the voice). Although such erroneous combinations of sensory cues can result in entertaining phenomena, they can also have dangerous consequences – for example, when combining the sound of one approaching car with the sight of another. How do we know which cues to integrate and which to ignore? One way would be to ignore (filter out) sensory cues that are irrelevant to the task at hand. For example, in the example above, we ought to ignore a car parked on the other side of the road when localizing the sound of the approaching car. But how does this selective ability develop, and how does it influence cue combination across the lifespan?

Based on infant studies we know that soon after birth babies can detect and learn cross-modal spatial correspondences (Bremner et al., 2011). For example, Morrongiello, Fenwick, and Chance (1998) showed that newborn babies expected audiovisual cues to remain collocated in space even when the cues were transposed to a different location. Similarly, many studies have demonstrated that young infants are more sensitive to changes in stimuli that occur concurrently in different sensory modalities than to changes in only one modality (e.g., Bahrack, 1992; Bahrack & Lickliter, 2004; Lewkowicz, 1996, 2000). For example, Bahrack (1992) found that

infants as young as three-and-a-half months were more sensitive to temporal changes of an audiovisual stimulus, than to temporal changes in an auditory or visual stimulus alone. In other words, as formulated in the intersensory redundancy hypothesis (e.g., Bahrick & Lickliter, 2000, 2002), to learn sensory correspondences young infants appear to rely initially on stimulus properties that are not specific to one modality alone. Interestingly, however, not all amodal properties seem to have the same relevance for young infants (2 to 6 months of age), with redundancy in temporal characteristics prevailing over that in spatial characteristics (Morrongiello, Fenwick, & Nutley, 1998). This greater reliance on audiovisual synchronization over spatial collocation is still present in adults, although it is greatly reduced (Lewald & Guski, 2003). Consistent findings in support of the early ability of humans to use redundant multisensory information come from neurophysiological studies with nonhuman animals (Wallace & Stein, 1997; Wallace, Wilkinson, & Stein, 1996), as well as from neurorecording studies with human infants (Hyde, Jones, Porter, & Flom, 2010; Reynolds, Bahrick, Lickliter, & Guy, 2014). For example, studies measuring single-cell responses in kittens (Wallace & Stein, 1997) and event-related potentials in 3-month-old infants (Hyde et al., 2010) have shown increased neural responses for simultaneous audiovisual stimulation compared to either visual or auditory stimulation alone.

Beyond infancy, multisensory processing continues to mature during childhood. For example, recent studies have shown that children are less sensitive to audiovisual asynchrony than adults (Hillock, Powers, & Wallace, 2011; Hillock-Dunn & Wallace, 2012). Furthermore, several independent studies using a Bayesian ideal observer framework have shown that young children, unlike adults, fail to combine multimodal

information in order to improve their precision (Gori, Del Viva, Sandini, & Burr, 2008; Gori, Giuliana, Sandini, & Burr, 2012; Gori, Sandini, & Burr, 2012; Nardini, Bedford, & Mareschal, 2010; Nardini, Begus, & Mareschal, 2012; Nardini, Jones, Bedford, & Braddick, 2008; Petrini, Remark, Smith, & Nardini, 2014). For example, while adults can optimally combine auditory and visual information to improve their precision when performing a spatial discrimination task (Alais & Burr, 2004; Gori, Sandini, et al., 2012), children younger than 12 years of age rely predominantly on visual information (Gori, Sandini, et al., 2012). However, this visual dominance appears to be age- and task-dependent, since studies examining infants' predictions based on sensory priming (Robinson & Sloutsky, 2004), and *children's* temporal discriminations (Gori, Sandini, et al., 2012), report a converse auditory dominance.

Reconciling the high sensitivity to intersensory redundancy found in infants (Bahrick, 1992; Bahrick & Lickliter, 2004) with the lack of optimal integration in young children (which in certain instances present itself as complete reliance upon one sense, e.g., Gori et al., 2008; Gori, Sandini, et al., 2012) is difficult due to a lack of studies exploring this issue. However, a first step towards a resolution would be to assume that different aspects of cue combination (intersensory redundancy and optimal integration) emerge and take precedence at different ages. That is, if young children (like infants) still rely strongly on temporal intersensory redundancy, then the presence of an irrelevant, but synchronous, visual event should bias their judgement of sound discrimination. This in turn would result in children exhibiting *suboptimal* integration (due to children attributing too much weight to the irrelevant but synchronous information). Such an account would also fit well with the recalibration theory proposed by Gori and colleagues (e.g., Gori et al., 2008; Gori, Sandini,

Martinoli, & Burr, 2010), which states that visual information may be used to calibrate other modalities for spatial tasks (e.g., discriminating objects' orientation) during early childhood. That is, if young children do use visual information to calibrate other sensory cues in spatial tasks, then they would not be able to discount or ignore visual information even when it is irrelevant.

Evidence showing that the ability to filter out irrelevant information increases with age has been limited so far to within-modality studies using flankers or auditory noise (Ridderinkhof, van der Molen, Band, & Bashore, 1997; Shepp & Barrett, 1991; Tipper, Bourque, Anderson, & Brehaut, 1989, Jones, Moore, & Amitay, 2015). Here we examine how selective the integration of multisensory information is during childhood, by testing whether 7--10 year old children and adults could ignore, and so avoid integrating, irrelevant but synchronized multisensory information. We investigated this using an auditory spatial discrimination task with or without a spatially displaced but temporally synchronous task-irrelevant visual cue. We hypothesized that if children's suboptimal integration does not depend at all on their inability to ignore irrelevant synchronous information, then they should perform well on such a task (i.e., they would attribute little or no weight to the irrelevant information). In contrast, if children's suboptimal integration does in part depend on their inability to ignore the irrelevant synchronous information, then they should perform badly on such a task (i.e., they would *over-weight* the irrelevant information).

Method

Participants

Seventeen children (7 to 10 years; 9 females and 8 males), and 15 adults (19 to 38 years; 9 females and 6 males) with normal sight and hearing participated. Participants, recruited by leaflets and press advertisements, were from a socially and ethnically diverse population in central London, UK. The data were collected from June 2012 to September 2013. Sample size was selected based on previous studies using similar paradigms that have fitted psychometric functions to children's data (i.e., using a high number of trials with children; see for example Barutchu et al., 2010; Gori et al., 2008; Gori, Sandini, et al., 2012; Nardini et al., 2008). Adults and children's parents or guardians gave informed consent to participate, and the study received ethical approval from the research ethics board of University College London.

Procedure

Figure 1 shows the experimental set up. Participants were seated comfortably in front of a computer screen at a distance of 1.7 meters. Nine loudspeakers were positioned along the azimuth, at a constant distance from both the participant (1.7 meters) and each other (13 cm). On each trial participants fixated a cross (2.5 x 2.5 cm) at the centre of the screen, and were then presented with two sounds (1000 Hz pure tones of 75 dB SPL, lasting 100 ms). The *standard* sound was always played through the speaker positioned in the middle (i.e., the fifth speaker, see Figure 1).

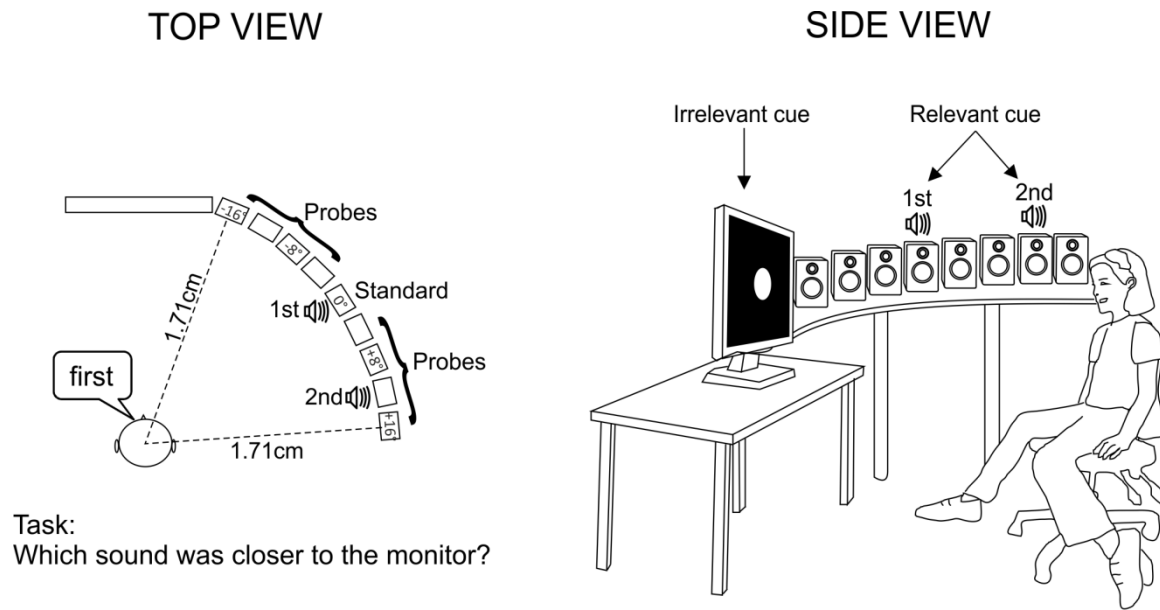


Figure 1. Experimental set-up. Nine speakers were positioned at equal distances to the observer, on the right side of a computer screen in a semicircle. The comparison sound could be presented either with or without a synchronous flash (the irrelevant cue), presented directly in front of the observer. The standard sound was always played by the speaker positioned in the middle (see plan view representation on the left). The comparison sound (probe) could be played by any of the nine speakers. The observer judged which sound (the standard or the comparison) was closer to the monitor, either in a noiseless environment or one in which all speakers played added auditory pink noise.

The *comparison* sound was played through any one of the nine speakers and its presentation was counterbalanced across trials. The standard and comparison sound were separated by 300 ms and their order of presentation was counterbalanced in pseudorandom order across the whole experiment (i.e., in half of the trials the standard was played first and then after 300 ms the comparison was played, in the other half of the trials the opposite occurred). The experiment consisted of two blocks (noiseless and noisy) of 180 trials, for a total of 360 trials. Within each block, trials were uniformly distributed across nine sound locations (positions of the speakers) and two visual conditions (presence and absence of the flash), and these conditions were

randomly interleaved. During the noisy block, both standard and comparison sound were combined additively with a random sample of pink noise. Pink noise is a random noise similar to the white noise (e.g., the ‘static’ caused by a detuned radio), but more pleasant to listen to due to the attenuation of high-frequency components. The pink noise was played from all nine speakers and had an intensity equal to 1/10 of either the standard and comparison sound. The standard and comparison sounds were identical except for spatial location, as shown in Figure 1. On each trial, the pink noise onset was 300 ms before the onset of the first sound and ended after the second sound. The noisy auditory condition was included for two reasons: 1) to exclude the possibility that age-related differences in the use of the visual irrelevant cue depended solely on differences in sound discrimination threshold; 2) to examine if the use of the visual irrelevant cue increased with the decrease in reliability of the relevant cue (less discriminable sound). The two visual conditions consisted of one in which both the standard and comparison sound were presented alone, and another in which the standard sound was presented alone while the comparison sound was presented together with a synchronous white flash (17 cm in diameter, 100 ms in duration) in the centre of the screen. In each trial the single flash was perfectly synchronised with the sound in onset, offset and duration. Synchronization in sound and flash presentation was achieved using a Focusrite Saffire PRO 40 sound card in conjunction with the Matlab Psychtoolbox ‘PsychPortAudio’ ASIO interface (Brainard, 1997; Pelli, 1997).

We asked participants to report verbally which of the two sounds was closer to the monitor. Participants underwent short practice runs separately for the noiseless and noisy conditions. Practice consisted of a few repetitions (a minimum of four, further

repeated if needed) of the easiest spatial discrimination trials – those using the first or ninth speaker; Figure 1). This initial practice made sure that the task was clear to participants, and that they had experience of the flash as an irrelevant stimulus.

Results

To test whether children, in contrast to adults, would be unable to ignore irrelevant visual information, we asked 7- to 10-year-old children and adults to perform a sound discrimination task, with or without an irrelevant visual flash presented synchronously in front of them. Participants judged whether the first or second tone in a pair was closer to the monitor. In each trial one tone was the standard sound (always played by the speaker positioned in the middle), while the other was the comparison sound (played by any one of the nine speakers). The proportion of trials in which the comparison sound was judged to be closer to the monitor than the standard (0 on the abscissa in Figure 2) was fitted with a cumulative Gaussian independently for each individual. Figure 2a and 2b (left panels) present the psychometric fits for a representative child and adult participant, obtained using *psignifit* 2.5.6 (see <http://bootstrap-software.org/psignifit/>), a software package that implements maximum-likelihood parameter estimation. The point at which the psychometric function intersects the 50% correct point on the ordinate is the point of subjective equality (PSE). The slope parameter of the functions provides a measure of the spatial discrimination sensitivity (steeper slope = more sensitive). As an exclusion criterion we used the sound only threshold in both noiseless and noisy conditions (i.e., as that was the parameter of interest, used to determine the relative weight observers gave to auditory and visual information, see Weight Analysis, below). The data of two children and two adults were excluded as their performance was at chance level in

either sound only (noiseless) or sound only with pink noise (noisy) conditions.

Threshold analysis

We first examined the effect of age on spatial discrimination thresholds for auditory-only stimuli in the two auditory noise conditions (no noise, pink noise). Figure 2c (left panel) shows that thresholds were higher for children than for adults, and higher with noise than without. A mixed model factorial ANOVA with age as a between-subjects factor and noise condition as within-subjects factor revealed significant main effects of age, $F_{(1, 30)} = 66.217, p < 0.001$, and noise condition, $F_{(1, 30)} = 20.847, p < 0.001$, and a significant interaction, $F_{(1, 30)} = 9.141, p = 0.005$. Paired-samples t -tests showed that children had significantly lower thresholds in the noiseless than the noisy condition, $t_{16} = -4.404, p < 0.001$, while adults did not differ between conditions, $t_{14} = -1.835, p = 0.088$. Independent-samples t -tests showed that children had significantly higher thresholds than adults under both noiseless, $t_{30} = 4.901, p < 0.001$, and noisy, $t_{30} = 8.192, p < 0.001$, conditions.

Weight analysis

We next examined the effect of age and noise level (without added noise and with added pink noise) on the relative weight attributed to the auditory cue (relevant cue) during the sound+flash condition (see the Supplemental Material for details of how the weights were calculated and for the Matlab code). Figure 2c (right panel) indicates that children weighted the irrelevant visual cue nearly as much as the relevant auditory cue (mean auditory weights just over 0.5). Conversely, adults largely ignored the visual cue (mean auditory weights close to 1). A mixed model factorial ANOVA with age as a between-subjects factor and noise condition as within-subjects factor

revealed a significant main effect of age, $F_{(1,30)} = 15.946$, $p < 0.001$, no main effect of noise condition, $F_{(1,30)} = 0.372$, $p = 0.547$, and no interaction, $F_{(1,30)} < 0.001$, $p = 0.985$.

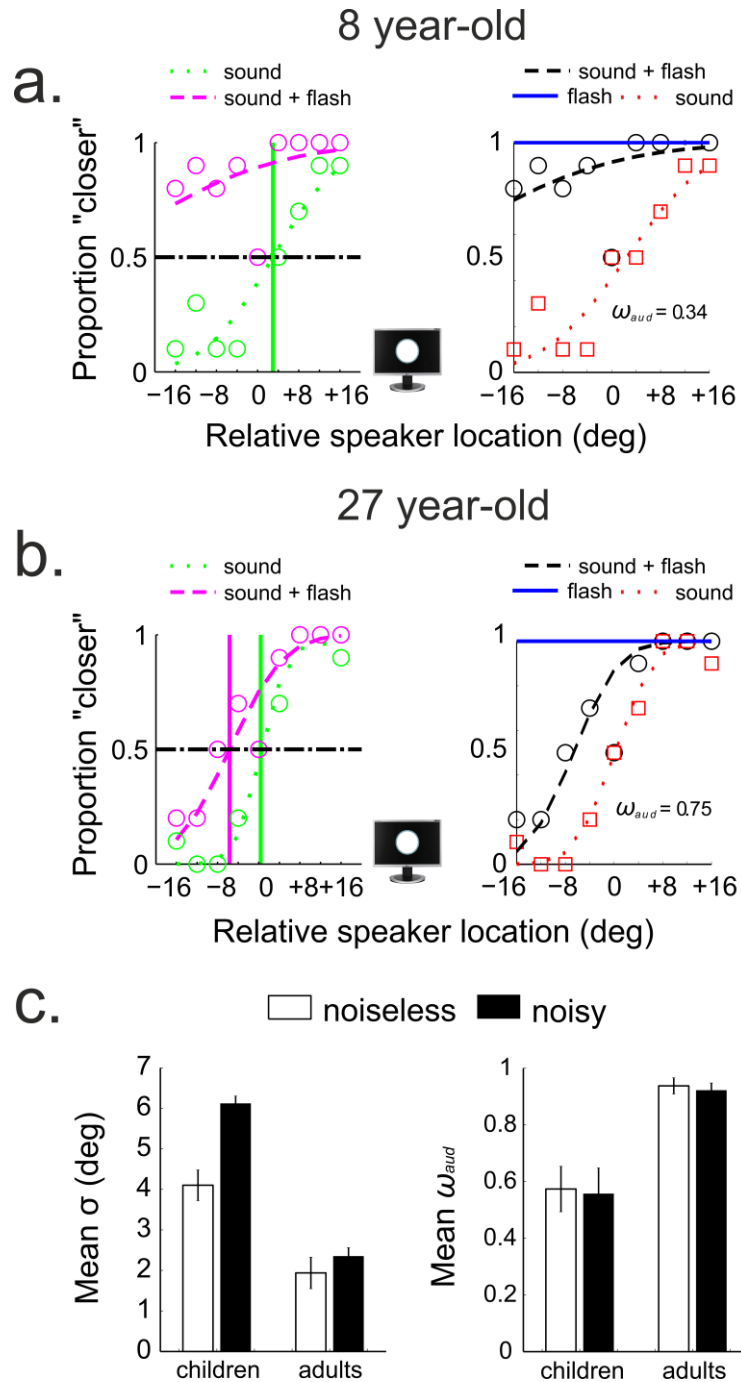


Figure 2. (a). Data from a child in the noiseless condition (1st in Figure 3, top panel). (b). Data from an adult in the noiseless condition (14th adult in Figure 3, bottom panel). Diagrams plot the proportion of ‘closer’ responses given to the comparison sound for two experimental conditions (dashed = sound +

flash; dotted = sound only). The monitor image indicates the screen position with respect to the speakers. Solid vertical lines in the left panels represent the point of subjective equality (PSE) obtained by fitting a cumulative Gaussian function to the data (dashed and dotted lines). For the child, this point fell outside the range tested for the sound + flash condition. The diagrams on the right show the same data, but here the sound+flash is fitted by a model (see equation 1 in the supplemental material) estimating the observer's auditory weight ω_{aud} , which determines where this curve lies between those predicted by use of vision only (solid horizontal line on top) vs. sound only (dotted line). The child in Fig 2a (right) is best fitted with an auditory weight of 0.34 (i.e., the dashed sound+flash line is closer to the solid flash line), the adult in Fig 2b (right) with an auditory weight of 0.75 (i.e., the dashed sound+flash line is closer to the dotted sound line). Since weights must sum to 1, the visual weight for each participant is $1-\omega_{\text{aud}}$. **(c)**. The diagram on the left plots children's and adults' mean thresholds for the sound only condition (obtained from the slope of the dotted cumulative function) in the auditory noiseless and noisy condition. The diagram on the right plots children's and adults' mean auditory weight (i.e. the weight attributed to the relevant cue) in the auditory noiseless and noisy condition, respectively. The error bars represent the standard error of the mean.

The mean weight attributed by children to the auditory cue (the relevant cue) during the sound+flash condition was not significantly different from 0.5 in either the noiseless, $t_{16} = 0.917$, $p = 0.373$, or noisy condition, $t_{16} = 0.619$, $p = 0.545$. In contrast, the mean weight attributed by adults was significantly different from 0.5 in both noiseless, $t_{14} = 15.358$, $p < 0.001$, and noisy, $t_{14} = 15.970$, $p < 0.001$. These results indicate that children's greater reliance on irrelevant visual information was not a consequence of their poorer auditory-only threshold. If children's high weighting for vision was explained by a lack of reliability in the auditory cue, then they should have weighted the visual cue even more during the noisy condition. However, although the noisy condition gave higher auditory-only thresholds (Fig 2c, left), children did not

weight vision differently in this condition as compared with the noiseless condition (Fig 2c, right).

Correlation analysis

We used two-tailed Pearson's correlation analyses to ask whether individual children's auditory weights were predicted by their auditory thresholds: they were not, in either noiseless, $r = 0.044$, $p = 0.876$, or noisy, $r = -0.223$, $p = 0.424$, conditions. Adults had weightings close to ceiling, 1, and so did not have the dispersion needed for a correlation analysis (see Figure 3).

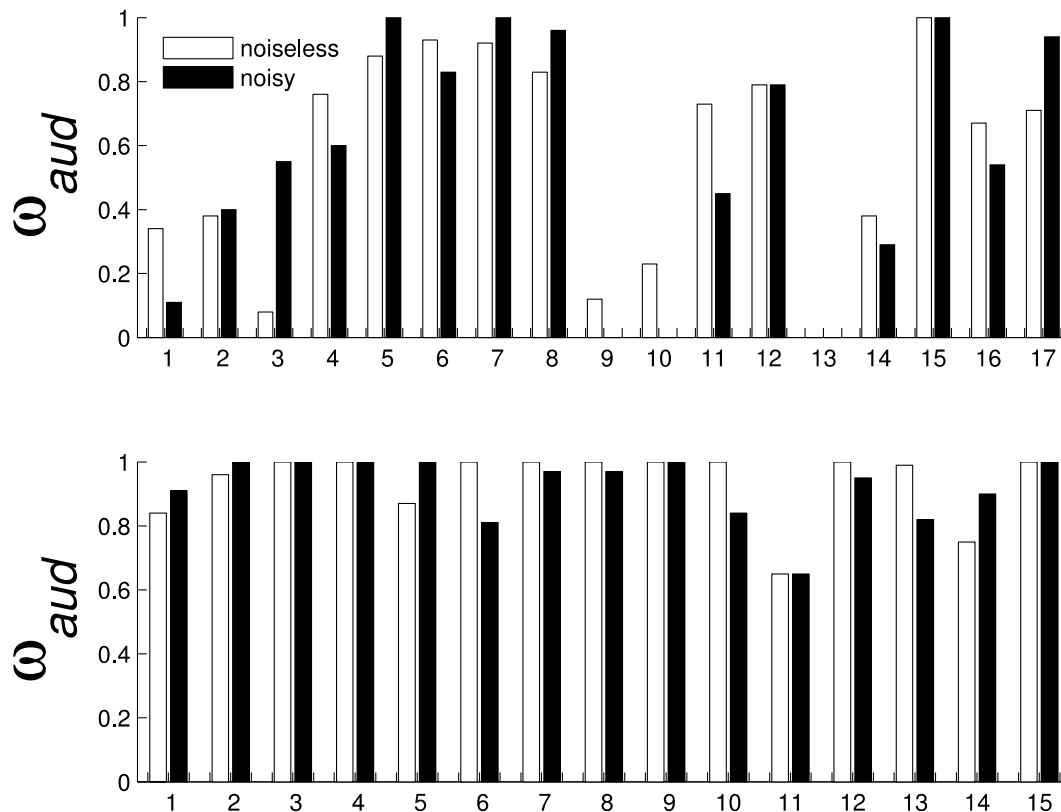


Figure 3. Weight attributed to the auditory cue in the sound+flash condition by each individual child (top panel) and adult (bottom panel) for noiseless (white bars) and noisy (black bars) blocks, respectively. The average data are plotted in Figure 2c right panel.

The results indicate that children's weighting of the irrelevant visual cue was not driven by the reliability of the auditory cue. The fact that children do not weight multimodal cues proportionally to their reliability is consistent with previous findings using task relevant cues (e.g. Gori et al., 2008; Nardini et al., 2008; Petrini et al., 2014), and with findings demonstrating that decreasing sound reliability by adding noise does not facilitate integration (Barutchu et al. (2010). Finally, a directional (one-tailed) Pearson's correlation analysis revealed that within the child group, the weight attributed to the relevant auditory cue increased with age, $r = 0.318$, $p = 0.034$. Bias corrected bootstrap 95% confidence intervals for r were [0.035, 0.580]. As weights sum to 1, this means that the relative weight given to the irrelevant cue decreased with age.

Discussion

The results indicated that children, in contrast to adults, weighted an irrelevant visual cue as much as a relevant auditory cue when discriminating between different locations of a sound. Hence, children performed poorly in a task that required ignoring irrelevant visual information. Furthermore, the weight attributed by children to the irrelevant visual cue decreased with age.

Our findings show that acquiring mature cue combination mechanisms is a multifaceted process, and that the late development of optimal cue combination for relevant information (e.g. Gori et al., 2008; Gori, Sandini, et al., 2012; Nardini et al., 2012; Nardini et al., 2008; Petrini et al., 2014) is paralleled by a late development of mechanisms selecting cues for combination. Being able to keep sensory information separate (Nardini et al., 2010) may be adaptive when the body is growing, as each

sense needs to be continuously recalibrated (Burr, Binda, & Gori, 2011). Our results support the recalibration theory by Gori and colleagues in that keeping cues separate could be adaptive for calibrating other sensory cues with vision, especially in spatial tasks (e.g., Burr et al., 2011; Gori et al., 2008; Gori et al., 2010).

In our task, young children relied on vision in a spatial localization task for which vision was not informative. Their over-weighting of irrelevant visual information demonstrates that children are poor multisensory selectors. Our findings of an immature multisensory selection mechanism in young children is in agreement with the findings of Innes-Brown et al. (2011) in the temporal domain. Innes-Brown et al. (2011) found that children were more susceptible than adults to the flash-beep illusion (Shams, Kamitani, & Shimojo, 2000), in that they were more affected by the number of beeps when judging the number of flashes. That is, they relied on auditory information in a visual task for which audition was not informative. In both these studies, children over-relied on the “dominant” modality – the one that would in normal circumstances be the most useful for the task: vision for spatial localization in the present study, and audition for temporal judgments (number of brief events) in Innes-Brown et al (2011).

These results can also be related to those reported for the McGurk effect (i.e., an illusion inducing the perception of a third new utterance 'da', by combining the video of a speaker pronouncing 'ga' with a superimposed sound of 'ba'; McGurk & MacDonald, 1976), for which children have been found to be less susceptible than adults (McGurk & MacDonald, 1976). Children’s lower susceptibility to this illusion indicates that they rely relatively more on audition and less on vision. In the McGurk

effect, audition is the “dominant” modality, in the sense that it would usually tend to be most useful for the task. Overall then, results from all three studies indicate that children’s cue selection is less flexible than that of adults, and is consistent with use of a “default” weighting that relies on cues that would normally be expected to be most useful. With development comes the ability to calibrate this reliance on different cues appropriately to specific sensory situations.

During infancy humans combine multiple senses by strongly relying on intersensory redundancy (Bahrick, 1992; Bahrick & Lickliter, 2002, 2004). Infants and later young children find synchronous redundant stimuli across visual and auditory modalities very salient. However, later in childhood multisensory processing continues to be refined in many ways before it becomes adult-like. For example, children become able to weight cues proportionally to their reliability (Gori et al., 2008; Gori, Sandini, et al., 2012; Nardini et al., 2012; Nardini et al., 2008; Petrini et al., 2014), and to increase their multisensory selectivity (Hillock et al., 2011; Hillock-Dunn & Wallace, 2012; Innes-Brown et al., 2011) to optimize precision. Here we show that despite the presence of individual differences among children and adults, which are very common in developmental literature on cue combination (e.g., Gori et al., 2008; Nardini et al., 2012; Petrini et al., 2014), children’s ability to ignore a visual irrelevant but temporally redundant information is much poorer. Our findings suggest that during early childhood there is a shift from associating cues based on simple heuristics, such as temporal co-occurrence (mechanism prioritized during infancy: Bahrick & Lickliter, 2000, 2002; Hyde et al., 2010; Lewkowicz, 1996, 2010), to combining cues using more complex integration models, and thus increasing the behavioral gain (Ernst & Banks, 2002) afforded by these previously learned

associations. Changes in levels of cue selectivity across the lifespan could be used as a marker to examine the progression across different stages of cue combination thus filling the currently existing gap in developmental knowledge.

To understand fully how multisensory mechanisms develop during childhood, we must study them in more complex and naturalistic settings where both relevant and irrelevant cues are present. This would also require the inclusion of an additional parameter – determining the extent to which cues are integrated or segregated – to the standard Bayesian model of cues combination (Ernst & Banks, 2002; Landy et al., 1995). The causal Bayesian inference model described by Kording et al. (2007) does take this new parameter into account to create an optimal Bayesian observer that not only infers the best sensory estimates from two sensory signals (i.e., inferring the source location from two sensory signals) but also whether the signals have a common cause. This kind of model may be tested in future to account for the differences between children and adults shown here.

At this stage we do not know whether the filtering of irrelevant sensory cues and the integration of the relevant cues is supported by the same brain mechanisms, and whether they share the same time courses and neural substrates during development. However, the finding that children's suboptimal integration may in part depend on their inability to ignore the irrelevant synchronous information suggests that there may be some overlap in brain processing. Building on recent studies examining the neural correlates of audiovisual speech processing (Reynolds et al., 2014) future neuroimaging and neurophysiological studies could help to answer these fundamental questions, by comparing children's and adults' neural activity in the presence and

absence of relevant and irrelevant cross-modal information. Knowing more about sensory selection and integration mechanisms will be of paramount importance for designing successful rehabilitation and treatment for congenital and acquired sensory deficits (Bebko, Weiss, Demark, & Gomez, 2006; Gogate, Maganti, & Perenyi, 2014; Lawson, Ruff, Mccartondaum, Kurtzberg, & Vaughan, 1984). For example, rehabilitation techniques using unimpaired senses to promote calibration during development (Gori, Sandini, Martinoli, & Burr, 2014) could be accompanied by rehabilitation aimed to help the development of multisensory selection mechanisms.

Acknowledgments

This work was supported by the James S. McDonnell Foundation 21st Century Science Scholar in Understanding Human Cognition Program, the NIHR Biomedical Research Centre for Ophthalmology at Moorfields and UCL, and the Nuffield Foundation Undergraduate Research Bursary Scheme. The authors would like also to thank the children and their parents or guardians for their participation.

References

- Alais, D., & Burr, D. (2004). The ventriloquist effect results from near-optimal bimodal integration. *Curr Biol*, *14*(3), 257-262. doi: 10.1016/j.cub.2004.01.029
- Bahrnick, L. E. (1992). Infants' perceptual differentiation of amodal and modality-specific audio-visual relations. *J Exp Child Psychol*, *53*(2), 180-199.
- Bahrnick, L. E., & Lickliter, R. (2000). Intersensory redundancy guides attentional selectivity and perceptual learning in infancy. *Dev Psychol*, *36*(2), 190-201.
- Bahrnick, L. E., & Lickliter, R. (2002). Intersensory redundancy guides early perceptual and cognitive development. *Adv Child Dev Behav*, *30*, 153-187.
- Bahrnick, L. E., & Lickliter, R. (2004). Infants' perception of rhythm and tempo in unimodal and multimodal stimulation: a developmental test of the intersensory redundancy hypothesis *Cogn Affect Behav Neurosci*, *4*(2), 137-147.
- Barutchu, A., Danaher, J., Crewther, S. G., Innes-Brown, H., Shivdasani, M. N., & Paolini, A. G. (2010). Audiovisual integration in noise by children and adults. *J Exp Child Psychol*, *105*(1-2), 38-50. doi: 10.1016/j.jecp.2009.08.005
- Bebko, J. M., Weiss, J. A., Demark, J. L., & Gomez, P. (2006). Discrimination of temporal synchrony in intermodal events by children with autism and children with developmental disabilities without autism. *J Child Psychol Psychiatry*, *47*(1), 88-98. doi: 10.1111/j.1469-7610.2005.01443.x
- Brainard, D. H. (1997). The Psychophysics Toolbox. *Spat Vis*, *10*(4), 433-436.
- Bremner, J. G., Slater, A. M., Johnson, S. P., Mason, U. C., Spring, J., & Bremner, M. E. (2011). Two- to eight-month-old infants' perception of dynamic auditory-visual spatial collocation *Child Dev*, *82*(4), 1210-1223. doi: 10.1111/j.1467-8624.2011.01593.x

- Burr, D., Binda, P., & Gori, M. (2011). Multisensory integration and calibration in adults and children. In K. K. Julia Trommershäuser, Michael S. Landy (Ed.), *Sensory cue integration* (pp. 173-194): Oxford University Press.
- Ernst, M. O., & Banks, M. S. (2002). Humans integrate visual and haptic information in a statistically optimal fashion. *Nature*, *415*(6870), 429-433. doi: 10.1038/415429a
- Gogate, L., Maganti, M., & Perenyi, A. (2014). Preterm and term infants' perception of temporally coordinated syllable-object pairings: implications for lexical development. *J Speech Lang Hear Res*, *57*(1), 187-198. doi: 10.1044/1092-4388(2013/12-0403)
- Gori, M., Del Viva, M., Sandini, G., & Burr, D. C. (2008). Young children do not integrate visual and haptic form information. *Curr Biol*, *18*(9), 694-698. doi: 10.1016/j.cub.2008.04.036
- Gori, M., Giuliana, L., Sandini, G., & Burr, D. (2012). Visual size perception and haptic calibration during development. *Dev Sci*, *15*(6), 854-862. doi: 10.1111/j.1467-7687.2012.2012.01183.x
- Gori, M., Sandini, G., & Burr, D. (2012). Development of visuo-auditory integration in space and time. *Front Integr Neurosci*, *6*, 77. doi: 10.3389/fnint.2012.00077
- Gori, M., Sandini, G., Martinoli, C., & Burr, D. (2010). Poor haptic orientation discrimination in nonsighted children may reflect disruption of cross-sensory calibration. *Curr Biol*, *20*(3), 223-225. doi: 10.1016/j.cub.2009.11.069
- Gori, M., Sandini, G., Martinoli, C., & Burr, D. C. (2014). Impairment of auditory spatial localization in congenitally blind human subjects. *Brain*, *137*(Pt 1), 288-293. doi: 10.1093/brain/awt311

- Hillock, A. R., Powers, A. R., & Wallace, M. T. (2011). Binding of sights and sounds: age-related changes in multisensory temporal processing. *Neuropsychologia*, *49*(3), 461-467. doi: 10.1016/j.neuropsychologia.2010.11.041
- Hillock-Dunn, A., & Wallace, M. T. (2012). Developmental changes in the multisensory temporal binding window persist into adolescence. *Dev Sci*, *15*(5), 688-696. doi: 10.1111/j.1467-7687.2012.01171.x
- Hyde, D. C., Jones, B. L., Porter, C. L., & Flom, R. (2010). Visual stimulation enhances auditory processing in 3-month-old infants and adults. *Dev Psychobiol*, *52*(2), 181-189. doi: 10.1002/dev.20417
- Innes-Brown, H., Barutchu, A., Shivdasani, M. N., Crewther, D. P., Grayden, D. B., & Paolini, A. G. (2011). Susceptibility to the flash-beep illusion is increased in children compared to adults. *Dev Sci*, *14*(5), 1089-1099. doi: 10.1111/j.1467-7687.2011.01059.x
- Jones, P. R., Moore, D. R., & Amitay, S. (2015). Development of auditory selective attention: Why children struggle to hear in noisy environments. *Developmental Psychology*, *51*(3), 353-369, doi: 10.1037/a0038570
- Kording, K. P., Beierholm, U., Ma, W. J., Quartz, S., Tenenbaum, J. B., & Shams, L. (2007). Causal inference in multisensory perception. *PLoS One*, *2*(9), e943. doi: 10.1371/journal.pone.0000943
- Landy, M. S., Maloney, L. T., Johnston, E. B., & Young, M. (1995). Measurement and modeling of depth cue combination: in defense of weak fusion. *Vision Res*, *35*(3), 389-412.
- Lawson, K. R., Ruff, H. A., Mccartondaum, C., Kurtzberg, D., & Vaughan, H. G. (1984). Auditory Visual Responsiveness in Full-Term and Preterm Infants. *Dev Psychol*, *20*(1), 120-127. doi: Doi 10.1037//0012-1649.20.1.120

- Lewald, J., & Guski, R. (2003). Cross-modal perceptual integration of spatially and temporally disparate auditory and visual stimuli. *Brain Res Cogn Brain Res*, *16*(3), 468-478.
- Lewkowicz, D. J. (1996). Perception of auditory-visual temporal synchrony in human infants. *J Exp Psychol Hum Percept Perform*, *22*(5), 1094-1106.
- Lewkowicz, D. J. (2000). The development of intersensory temporal perception: an epigenetic systems/limitations view. *Psychol Bull*, *126*(2), 281-308.
- Lewkowicz, D. J. (2010). Infant perception of audio-visual speech synchrony. *Dev Psychol*, *46*(1), 66-77. doi: 10.1037/a0015579
- McGurk, H., & MacDonald, J. (1976). Hearing lips and seeing voices. *Nature*, *264*(5588), 746-748.
- Morrongiello, B. A., Fenwick, K. D., & Chance, G. (1998). Crossmodal learning in newborn infants: Inferences about properties of auditory-visual events. *Infant Behavior & Development*, *21*, 543-554.
- Morrongiello, B. A., Fenwick, K. D., & Nutley, T. (1998). Developmental changes in associations between auditory-visual events. *Infant Behavior & Development*, *21*, 613-626.
- Nardini, M., Bedford, R., & Mareschal, D. (2010). Fusion of visual cues is not mandatory in children. *Proc Natl Acad Sci U S A*, *107*(39), 17041-17046. doi: 10.1073/pnas.1001699107
- Nardini, M., Begus, K., & Mareschal, D. (2012). Multisensory Uncertainty Reduction for Hand Localization in Children and Adults. *J Exp Psychol Hum Percept Perform*. doi: 10.1037/a0030719

- Nardini, M., Jones, P., Bedford, R., & Braddick, O. (2008). Development of cue integration in human navigation. *Curr Biol*, *18*(9), 689-693. doi: 10.1016/j.cub.2008.04.021
- Pelli, D. G. (1997). The VideoToolbox software for visual psychophysics: transforming numbers into movies. *Spat Vis*, *10*(4), 437-442.
- Petrini, K., Remark, A., Smith, L., & Nardini, M. (2014). When vision is not an option: children's integration of auditory and haptic information is suboptimal. *Developmental Science*, 1-12.
- Reynolds, G. D., Bahrick, L. E., Lickliter, R., & Guy, M. W. (2014). Neural correlates of intersensory processing in 5-month-old infants. *Dev Psychobiol*, *56*(3), 355-372. doi: 10.1002/dev.21104
- Ridderinkhof, K. R., van der Molen, M. W., Band, G. P., & Bashore, T. R. (1997). Sources of interference from irrelevant information: a developmental study. *J Exp Child Psychol*, *65*(3), 315-341. doi: 10.1006/jecp.1997.2367
- Robinson, C. W., & Sloutsky, V. M. (2004). Auditory dominance and its change in the course of development. *Child Dev*, *75*(5), 1387-1401. doi: 10.1111/j.1467-8624.2004.00747.x
- Shams, L., Kamitani, Y., & Shimojo, S. (2000). Illusions. What you see is what you hear. *Nature*, *408*(6814), 788. doi: 10.1038/35048669
- Shepp, B. E., & Barrett, S. E. (1991). The development of perceived structure and attention: evidence from divided and selective attention tasks. *J Exp Child Psychol*, *51*(3), 434-458.
- Tipper, S. P., Bourque, T. A., Anderson, S. H., & Brehaut, J. C. (1989). Mechanisms of attention: a developmental study. *J Exp Child Psychol*, *48*(3), 353-378.

Wallace, M. T., & Stein, B. E. (1997). Development of multisensory neurons and multisensory integration in cat superior colliculus. *J Neurosci*, *17*(7), 2429-2444.

Wallace, M. T., Wilkinson, L. K., & Stein, B. E. (1996). Representation and integration of multiple sensory inputs in primate superior colliculus. *J Neurophysiol*, *76*(2), 1246-1266.