PAPER

When vision is not an option: children's integration of auditory and haptic information is suboptimal

Karin Petrini,[1] Alicia Remark,[1] Louise Smith[2] and Marko Nardini[1]

1. Institute of Ophthalmology, University College London, UK
2. Department of Experimental Psychology, University of Oxford, UK

Abstract

When visual information is available, human adults, but not children, have been shown to reduce sensory uncertainty by taking a weighted average of sensory cues. In the absence of reliable visual information (e.g. extremely dark environment, visual disorders), the use of other information is vital. Here we ask how humans combine haptic and auditory information from childhood. In the first experiment, adults and children aged 5 to 11 years judged the relative sizes of two objects in auditory, haptic, and non-conflicting bimodal conditions. In Experiment 2, different groups of adults and children were tested in non-conflicting and conflicting bimodal conditions. In Experiment 1, adults reduced sensory uncertainty by integrating the cues optimally, while children did not. In Experiment 2, adults and children used similar weighting strategies to solve audio–haptic conflict. These results suggest that, in the absence of visual information, optimal integration of cues for discrimination of object size develops late in childhood.

Research highlights

- Children and adults were tested on the ability to integrate haptic and auditory information about object size, without vision.
- Optimal integration of sensory estimates in the absence of visual information occurs late.
- Adults and children use similar weighting strategies to solve audio–haptic conflict.
- Pre-adolescents specifically lack the ability to reduce the variability of their responses through appropriately weighted averaging.

Introduction

It has been shown that adults can combine a visual sensory estimate with other sensory estimates (either visual or non-visual) to reduce sensory uncertainty when judging, for example, the size, shape or the position of an object (e.g. Alais & Burr, 2004; Ernst & Banks, 2002; Hillis, Ernst, Banks & Landy, 2002). In these tasks, human adults optimally combined their sensory estimates, reducing their variance in line with a Bayesian ideal observer. Recent studies have demonstrated that, in contrast to human adults, children as old as 8–10 years are not yet able to optimally combine the kind of sensory information listed above. For example, Gori, Del Viva, Sandini and Burr (2008) showed that children younger than 8–10 years did not optimally integrate visual and haptic information to reduce uncertainty when discriminating object size or orientation. Instead, they gave too much weight to the sense which was less reliable for the task. Similarly, Nardini, Jones, Bedford and Braddick (2008) showed that young children did not integrate visual and movement-related information during spatial navigation. Children under 12 kept two visual cues relating to the angle of a surface separate (Nardini, Bedford & Mareschal, 2010), not reducing uncertainty by combining them. This allowed them to avoid the 'sensory fusion' experienced by adults with conflicting stimuli (Hillis *et al.*, 2002). This developmental trend for late maturation of integration mechanisms has also been extended to audiovisual integration, which appears to remain immature until at least 10–11 years of age

(Barutchu, Crewther & Crewther, 2009; Barutchu, Danaher, Crewther, Innes-Brown, Shivdasani & Paolini, 2010; Innes-Brown, Barutchu, Shivdasani, Crewther, Grayden & Paolini, 2011).

The reason why optimal use of multisensory information to reduce the uncertainty of estimates develops so late is yet to be determined, but different possibilities have been suggested. One possibility is that during childhood, different sensory estimates need to be kept separate so that the sensory system can be continuously recalibrated (Gori, Sandini, Martinoli & Burr, 2010; Gori, Tinelli, Sandini, Cioni & Burr, 2012). In children, sensory calibration is more important than achieving more precise estimates through integration. Depending on the task, the sense that is used as the benchmark for recalibration changes (Gori *et al.*, 2008; Gori *et al.*, 2010; Gori *et al.*, 2012). Another possibility is that during development the sensory system may be optimized for speed over accuracy, and so might use the fastest available single estimate (Nardini *et al.*, 2010).

Both possibilities share the common idea that, with age, the sensory system varies the importance of uncertainty reduction as compared with other goals. Unsurprisingly, all of the developmental evidence so far has come from studies in which one of the sensory information sources under study was vision, as for healthy humans vision is a key component of most naturalistic tasks. However, certain environments (e.g. an extremely dark surrounding) and visual disorders can reduce or even eliminate the visual input. In these cases, the use of other information, such as touch and sound, is vital.

Here we aimed to extend what we know about the development of uncertainty reduction by using a non-visual task. Studies examining the interaction between haptic and auditory cues as well as neural substrates of this interaction in adults (e.g. Bresciani & Ernst, 2007; Kassuba, Menz, Roder & Siebner, 2012; Sanabria, Soto-Faraco & Spence, 2005; Soto-Faraco & Deco, 2009) point to a common mechanism for reducing sensory uncertainty in the presence or absence of visual information. Because the interaction between haptic and auditory cues has not been examined in children, it is yet to be determined whether multisensory integration of non-visual information develops at a similar age as that involving vision.

In the real world, the likelihood of a causal relationship between haptic and auditory information is often higher than, for example, that between visual and haptic or audio and visual information. This is because haptic-auditory conjunctions (e.g. touching an object against a surface) tend to be produced by our own actions (Tajadura-Jimenez, Valjamae, Toshima, Kimura, Tsakiris & Kitagawa, 2012), while visual-auditory conjunctions are often caused by external events. In a visual-haptic conjunction (e.g. looking at an object while touching an object) the two sensory inputs are only related when the same object is being looked at and touched, which need not be the case. In a haptic-auditory situation like the present one, the participant's own action causes the sound. Thus, at least in otherwise quiet environments, there is a very strong basis for linking the haptic and auditory information. It is therefore possible that the kind of auditory-haptic integration tested in the present study would develop earlier than visual-haptic integration as tested previously. Alternatively, audio-haptic integration could show even later development than visual-haptic because vision is a very dominant sense in other audiovisual tasks (Nava & Pavani, 2013) during the early school years tested in the present study. Both of these outcomes would imply separate and patchy maturation of multisensory mechanisms during childhood. If the development of visual-haptic and audio-haptic integration occurs at the same time, this would imply a common and single multisensory mechanism in place from early childhood.

To test these possibilities we investigated the time course of haptic-auditory integration for size discrimination in the absence of visual information. In Experiment 1 we measured haptic-auditory integration by assessing size discrimination in auditory, haptic and non-conflicting haptic-auditory conditions in one group of 5–6-, 7–8-, 10–11-year-old children and adults. This allowed us to examine at what age sensory uncertainty was reduced by integrating the cues optimally (i.e. as predicted by the maximum likelihood estimation model). In Experiment 2 we further examined the development of haptic-auditory integration by examining size discrimination in haptic-auditory conditions with three different levels of cue conflict in a second group of 5–6-, 7–8-, 10–11-year-old children and adults. This allowed us to examine the relative weighting (reliance on) haptic and auditory information about size across ages.

Method

Auditory stimuli selection

Both changes in pitch (especially as a consequence of changes in fundamental frequency) and in amplitude and thus loudness are important for judgments of the size of objects striking a surface (Grassi, 2005). For this reason, before deciding on the best auditory stimuli to use in our study we ran pilot experiments with adult participants. We ran two experiments, one with changes only in pitch, and one with changes only in amplitude/loudness.

Participants judged which one of two wooden balls (a standard and a comparison ball) was bigger based either on touch, sound, or both (please refer to the *Procedure, stimuli, and design* section for details on task and loudness manipulation). For the pitch manipulation, the fundamental frequency (F0) of the middle-range ball (i.e. the standard ball with a F0 of 151 Hz) dropping on a wooden surface was increased or decreased by 15 Hz for the other eight comparison stimuli, giving rise to nine sounds ranging in F0 from 211 to 91 Hz. Psychometric functions were fitted to the proportion of 'bigger' responses given by each participant as a function of comparison stimulus size (please refer to the *Results* section for details on fitting and exclusion criteria). Strikingly, in the pitch experiment, 45% (5 of 11) of adults were unable to use pitch differences reliably to discriminate size. Their psychometric functions could not be fitted and they had to be excluded from analysis. By contrast in the amplitude experiment, 86% (12 of 14) used amplitude reliably.

Their ability to base size judgments on amplitude is in line with and supported by previous work in which adults judged the size of balls hitting a resonating plate (Grassi, 2005). The plate's oscillation increases with the mass of the object hitting it. Consequently, for larger objects the amplitude of the acoustic waveform is greater, and so is the loudness of the resulting sound. The results from the Grassi (2005) study show a strong correlation between amplitude domain indexes (power in particular) and the size of the ball. Indeed, the amplitude (power in particular), which is directly related and proportionate to perceived loudness, was the strongest predictor of participants' judgments of ball size. Similarly, in our study, the standard sound was recorded hitting a resonating surface (wooden table) after being dropped from a standard height (150 mm; Grassi, 2005). For participants the task was set up as if they were producing sounds by patting the balls against a similar table. So it is not surprising that pitch was not used as a reliable cue, as in this situation most of the sound is produced by the surface (table top), which does not change in pitch with object size. The poor ability of adults in our pilot studies to use pitch convinced us that we should use amplitude, in order to make the task feasible for children when using our set-up. It could also be possible, however, that the use of pitch for size discrimination decreases with age and that our pilot results for the pitch condition reflect adults' decreased ability to use the pitch information. Although we cannot completely exclude this possibility, Grassi (2005) showed that when the size of the resonating surface was changed (changing the pitch of the sound made by the balls) adults did use pitch information when judging the sizes of balls. Furthermore, developmental studies have shown that although both children and adults can use pitch information in several other kinds of perceptual task (e.g. music perception and speaker recognition; Demorest, 1992; Demorest & Serlin, 1997; Petrini & Tagliapietra, 2008), this ability improves with age (i.e. adults use the pitch information more than children when, for example, discriminating between speakers).

We also chose to vary one stimulus dimension – amplitude – to create the stimuli, rather than use naturally recorded sounds of the different sized balls. The rationale for this was to be able to conclude which cue all participants were using. If more than one cue is available, it is possible for developmental differences to reflect differences in strategy or attention to different cues.

Procedure, stimuli, and design

Participants sat in a comfortable chair behind a black curtain that covered the experimental set-up and stimuli (Figure 1). All participants were then asked to slide the dominant hand (as assessed using the Oldfield Edinburgh Handeness Inventory) through a hole in the curtain and rest their arm on a semi-soft foam surface. In the middle of the rectangular foam surface was a square
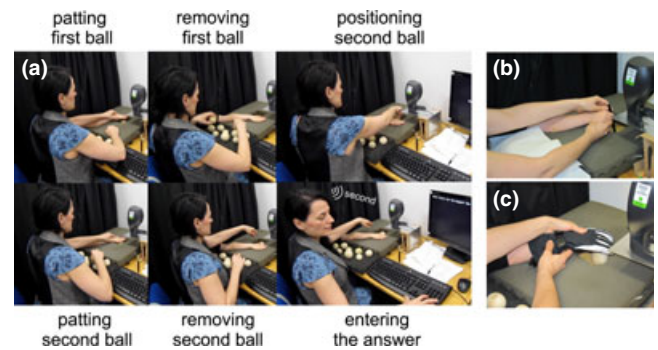
Figure 1 *Experimental set-up. (a) Experimental set-up and trial description. Participants sat behind a black curtain with their dominant hand inserted through the curtain and positioned comfortably on a semi-soft foam surface. A speaker was positioned as close as possible to the position at which the hand of the participant would pat either the wooden ball or a pen on the surface underneath. A touchscreen was placed just below the foam and, when pressed, the sound was played through the speaker. (b) An example of the auditory-only condition during which participants tapped a pen on the touch screen twice to hear the sound made by each ball, before judging which sound was produced by the bigger ball. (c) An example of the haptic condition during which participants patted a pair of wooden balls (one at the time) while wearing a thick glove, before judging which ball was bigger.*

hole, 5 cm deep, into which the stimuli were positioned one at a time. The stimuli consisted of nine wooden balls, differing in diameter by 2 mm (range 41–57 mm). The sound of the standard ball (49 mm) hitting a wood surface was recorded with a D7 LTD dynamic microphone through Focusrite Saffire PRO 40 sound card using the Psychtoolbox PsychPortAudio command library (Brainard, 1997; Pelli, 1997). We recorded the sound by dropping the ball from a standard height. If we recorded the sound by hitting the ball then the size of the hand at the recording time (for example that of the experimenter) would influence the force of the impact and the resulting sound. The sound thus recorded would have been more consistent with adults' than children's expectations of what the resulting sound should be. We wished to avoid this, especially when testing children of different ages and adults for whom hand size and strength vary greatly. Loudness was increased or decreased by 1 dB for the other eight stimuli (following piloting), giving rise to nine sounds ranging in amplitude from 71 to 79 dB.

During the experiment, balls were placed in the hole one at a time by the experimenter, and the participant's hand was placed on top of the ball. On each trial, participants were asked to touch two different wooden balls in sequence, keeping their hand flat (i.e. they did not grasp the ball but only patted it). We asked participants to pat the ball rather than grasp it, because similarly to the auditory case we wanted to vary one stimulus dimension – height. If participants could grasp the ball then they could use its weight as well, and, as for the auditory cue, we would not know whether developmental differences would reflect differences in strategy or attention to different cues. Another very thin (1 cm) layer of soft foam was inserted between the thick layer of foam and a touch screen positioned underneath (Figure 1) to eliminate any impact sound between the wooden balls and the hard surface of the touch screen. The ball's sound was played through a speaker, positioned as close as possible to the position of the ball (Figure 1b–c), when pressure was sensed on the touch screen. Synchronization between the pressure elicited on the touch screen and the played sound was achieved using a Focusrite Saffire PRO 40 sound card in conjunction with the Psychtoolbox PsychPortAudio command library (Brainard, 1997; Pelli, 1997). After patting the second ball participants reported which one of the two stimuli was the bigger (the first or the second). The standard stimulus was always present in the pair, but its position was unpredictably first or second, with equal numbers as first and second within each experimental condition. A one-back randomization was used, i.e. random orders were generated until

an order in which no stimulus appears twice in a row was found and the experimenter followed the instructions displayed on the computer screen to know which pair of stimuli to present. The number of trials/repetitions, the exclusion criteria, method of fitting and number of subjects are like those previously used in a similar visual-haptic size discrimination task (Gori *et al.*, 2008).

Experiment 1

Participants

A total of 34 children (eight 5- to 6-year-olds, 16 7- to 8-year-olds, and ten 10- to 11-year-olds) and 12 adults (aged between 19 and 35) participated in the study. This number does not include children who were excluded from the study because they performed at chance level in one or more task conditions (see below). In both experiments the adults and the children's parents or guardians gave informed consent for participation in the study, which received ethical approval from the research ethics board of University College London.

Procedure

Participants took part in one of two experimental conditions each consisting of three different trial types, the order of which was counterbalanced across participants. Participants were assigned iteratively to each experimental condition in order to minimize sampling issues. In some conditions, to reduce the haptic reliability we asked participants to wear thick skiing gloves which varied in size from extra-small to extra-large to fit as well as possible children's and adults' hand sizes. The purpose of the gloves was to reduce participants' sensitivity during patting, thus decreasing the quality of the haptic information. A group of participants performed a condition including auditory-only, haptic-only without glove, and bimodal without glove trials. During the auditory-only condition participants tapped a pen on the touch screen (Figure 1b) instead of the balls to elicit the sounds. During this condition the experimenter took back the pen from the participant after every first sound and gave it back before every second sound, matching the task and timing as closely as possible to the other conditions. A second group performed a condition including auditory-only, haptic-only with glove, bimodal with glove trials. Each participant completed three blocks of 54 trials (including each of the nine comparisons six times) in each of three conditions, for a total of 162 trials.

Experiment 2

Participants

A total of 22 children (six 5- to 6-year-olds, ten 7- to 8-year-olds, and six 10- to 11-year-olds) and six adults (aged between 19 and 35) participated in the study. This number does not include children who were excluded from the study because they performed at chance level in one or more task conditions.

Procedure

Participants performed a condition including a bimodal congruent condition without glove (the same as the bimodal condition without glove in the first group) and two bimodal incongruent conditions without glove. In these incongruent conditions the auditory and haptic cues provided by the standard stimulus were in conflict, indicating different sizes, but averaging to 49 mm similarly to the congruent bimodal standard

ball. In one bimodal incongruent condition the haptic size of the standard stimulus was 49 + 4 mm and that of the auditory standard stimulus 49−4 mm, in the other the haptic size was 49−4 mm and the auditory 49 + 4 mm. The comparison stimuli were the same as in Experiment 1.

Results

Cumulative Gaussian functions were fitted to the proportion of 'bigger' responses given by each participant as a function of comparison stimulus size. The estimate of each individual's function's mean (i.e. the point at which the psychometric function cuts the 50% of 'bigger' responses) indicated the Point of Subjective Equality (PSE). The size discrimination threshold was given by the standard deviation of the psychometric function (i.e. the slope of the function). The fit shown in Figure 2 for each age group was obtained by averaging the fit obtained from each individual. Overall, the
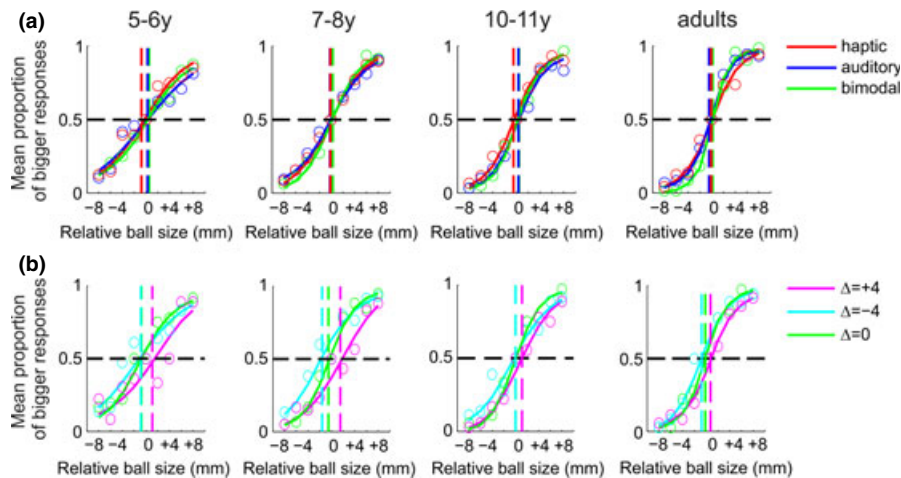
Figure 2 *Averaged psychometric functions for child and adult participants. The proportion of trials in which the comparison ball (whose size relative to the standard is given by the abscissa) was judged to be bigger than the standard (0 on the abscissa) was fitted with a cumulative Gaussian separately for each individual. For the fit we used psignifit version 2.5.6 (see http://bootstrap-software. org/psignifit/), a software package that implements the maximum likelihood method. Here to aid visualization we plot the fit results obtained by averaging the data of individuals within each age group. The point at which the psychometric function cuts the 50% point on the ordinate is the mean or PSE. The vertical dashed lines indicate the average PSEs. The slope of the functions is used to estimate the standard deviation or size discrimination threshold, such that the steeper the slope the lower is the variability and consequently the threshold. (a) Average results for the group of participants performing the haptic-only, auditory-only and bimodal congruent condition (i.e. no conflict between the cues). The red curve, symbols and dashed line refer to the average results for the haptic condition, the blue curve, symbols and dashed line to the auditory, and the green curve, symbols and dashed line to the congruent bimodal. (b) Average results for the group of participants performing one bimodal congruent and two bimodal incongruent conditions. The levels of cue conflict for the standard ball are represented here as −4, 0, and +4 mm for the haptic and +4, 0, and −4 for the auditory. A shift of the magenta dashed line toward +4 indicates that participants are relying more on the haptic information, whereas a shift toward −4 indicates that they are relying more on the auditory. The opposite is the case for the cyan line. The green curve, symbols and dashed line refer to the congruent bimodal condition (zero conflict between the cues), as in the same condition in (a).*

percentage of children excluded was 22% for 5–6-year-olds, 13% for 7–8-year-olds, and 15% for 10–11-year-olds, all of them due to an inability to do the haptic-only task when wearing the glove (i.e. they were not able to discriminate the standard ball from the others and their PSE or threshold fell outside the chosen range of stimuli). No adults had to be excluded for this reason.

Experiment 1

We tested the goodness of fit for each participant and found that the overall mean R^2 in the bimodal congruent condition was 0.87 $SE \pm 0.02$, for the haptic condition was 0.81 $SE \pm 0.02$, and for the auditory condition was 0.84 $SE \pm 0.01$, indicating that the chosen Gaussian psychometric function fitted the data well. We carried out a mixed factorial ANOVA with stimulus condition (auditory-only, haptic-only and bimodal congruent) as within-subjects factor and age group as between factor on the R^2 values. We found no significant interaction between age and condition ($F(6, 82) = 0.923$, $p = .483$), but a main effect of age ($F(3, 41) = 3.545$, $p = .023$). T-tests, Bonferroni corrected, post-hoc analyses showed that only the younger group of children differed significantly in their R^2 compared to adults ($p = .017$). Nevertheless, the mean R^2 of the 5–6-year-old children for the bimodal congruent condition was 0.75 $SE \pm 0.06$, i.e. the method of fitting still explained the young children's data well.

Before using parametric procedures to test differences in the thresholds obtained we tested whether the data could be approximated to a normal distribution. For this purpose we performed Kolmogorov–Smirnov tests of normality separately for each age group and experimental condition. We found that in 10 out of 12 cases the data were normally distributed (i.e. the tests of normality gave non-significant results), and in the two remaining cases (i.e. the adults' data in the auditory condition and the 10–11-year-old children's data in the bimodal condition) the data approximated normality ($p = .03$, $p = .021$). Overall, the thresholds obtained approximated well a normal distribution, thus allowing us to perform parametric analyses as previously done in similar studies (e.g. Gori *et al.*, 2008; Gori *et al.*, 2010; Nardini *et al.*, 2008).

An initial mixed-model ANOVA with noise level as between-subjects factor (no glove, glove) and stimulus type as within-subjects factor (auditory-only, haptic-only and bimodal) was carried out within each age group to examine the effect of noise on the size discrimination thresholds. This analysis revealed no main effect of noise level (5–6y: $F(1, 6) = 0.104$, $p = .758$; 7–8y: $F(1, 14) = 1.530$, $p = .236$; 10–11y: $F(1, 8) = 2.757$, $p = .135$; adults: $F(1, 10) = 2.580$, $p = .139$), and no interaction between

noise level and stimulus type (5–6y: $F(3, 18) = 0.544$, $p = .658$; 7–8y: $F(3, 42) = 0.582$, $p = .63$; 10–11y: $F(3, 24) = 2.104$, $p = .126$; adults: $F(3, 30) = 0.196$, $p = .898$). Based on this initial analysis we combined the data for the two noise level groups within each age group before further analysis.

To examine the effect of age on the size discrimination threshold for the different conditions, we carried out a mixed model factorial ANOVA with age as a between-subjects factor and condition as within-subjects factor. This analysis revealed a significant main effect of age ($F(3, 42) = 11.868$, $p < .001$), a main effect of condition ($F(3, 126) = 13.477$, $p < .001$), and a significant interaction between age and condition ($F(9, 126) = 1.911$, $p = .048$). T-tests, Bonferroni corrected, post-hoc analyses showed that overall adults had significantly lower thresholds than 5–6-year-old ($p < .001$), 7–8-year-old ($p < .001$), and 10–11-year-old children ($p = .017$). The older group of children had significantly lower thresholds compared to the 5–6-year-old children ($p = .016$), but not compared to the 7–8-year-old children ($p = .476$). Finally, 5–6- and 7–8-year-old children did not differ in their estimated thresholds ($p = 1$). Figure 3
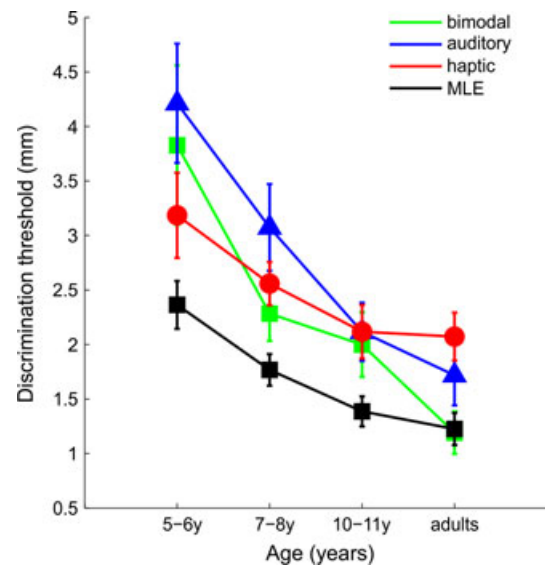
Figure 3 *Mean size discrimination thresholds as a function of age. Mean discrimination thresholds for haptic (red line and symbols), auditory (blue line and symbols), and bimodal congruent (green line and symbols) conditions as a function of age. The black line and symbols represent the average MLE model predictions for the bimodal condition as a function of age. The predicted bimodal threshold (σ_{HA}) was calculated individually for each subject, and then averaged, by entering the individual haptic (σ_H) and auditory (σ_A) thresholds into the equation $\sigma_{HA}^2 = \sigma_H^2 \sigma_A^2 / \sigma_H^2 + \sigma_A^2$. Error bars represent the standard error of the mean.*
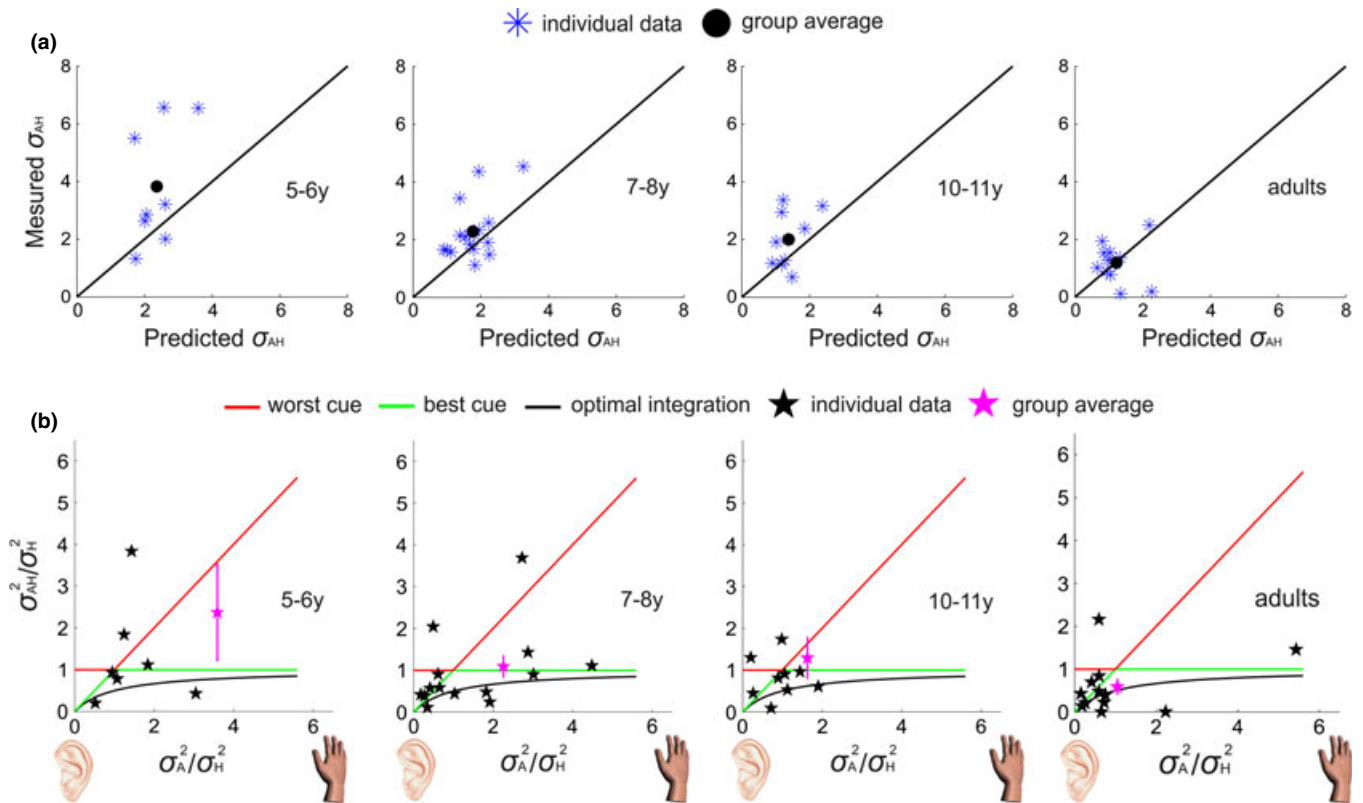
Figure 4 *Measured bimodal thresholds against MLE-predicted bimodal thresholds for individuals in each age group. (a) The blue symbols represent the individual data and the black dots indicate the group average thresholds. The observed thresholds come closer to the MLE prediction with age, as indicated by the black dot approaching the solid black line. (b) Ratios of single-cue variances (A/H) and combined-to-single-cue variances (AH/H) for individuals in each age group. Individual and group average performance is plotted together with predictions based on use of the single worst cue (red line), the single best cue (green line), or integration of cues according to the Bayesian model (black line). High ratios along the x-axis correspond to much more reliable touch than sound. Low ratios (< 1) along the y-axis corresponds to an improvement given both cues compared to touch alone. To aid visualization individual data out of the represented x and y range are not shown.*

summarizes these results and clearly shows that although discrimination thresholds decrease with age in all conditions, the bimodal discrimination thresholds were only well predicted by the optimal (ideal observer) estimate for adults. The optimal estimate was calculated by entering the unimodal discrimination thresholds into the maximum likelihood (MLE) model.

This observation was further supported by a series of planned one-tailed *t*-tests showing that the bimodal threshold for adults was significantly lower than both the haptic ($t(11)$ = −3.279, p = .007) and the auditory threshold ($t(11)$ = −1.428, p = .05), and did not differ from the predicted optimal bimodal threshold ($t(11)$ = −0.134, p = .44). This was not, however, the case for the children. The 10- to 11-year-old group's bimodal threshold was significantly higher than that predicted by MLE ($t(9)$ = 2.249, p = .02), and was no different from either the haptic ($t(9)$ = −0.422, p = .34) or the auditory

threshold ($t(9)$ = −0.400, p = .34). The 7- to 8-year-old group's bimodal threshold, similarly to the older children, was higher than that predicted by MLE ($t(15)$ = 2.389, p = .01), and was no different from the haptic ($t(15)$ = −0.958, p = .17), but was significantly lower than the auditory ($t(15)$ = −2.262, p = .01). Finally, the 5- to 6-year-old group's bimodal threshold was significantly higher than that predicted by MLE ($t(7)$ = 2.247, p = .02), but was no different from either the haptic ($t(7)$ = 0.866, p = .20) or the auditory threshold ($t(7)$ = −0.678, p = .26).

Figure 4a shows how the relationship between optimal predicted and measured bimodal discrimination threshold changes with age. It also shows how the distribution of individual performance becomes more consistent (less variable) with age, with a higher number of individuals performing as an ideal observer. Figure 4b plots the relationship between ratios of single-cue variances (A/H)

and combined- to single-cue variances (AH/H) at each age. The three lines indicate the predictions of different cue combination rules: relying on the worst single cue (red line), the best single cue (green line), or using the Bayesian model (black curve). Ratios greater than 1 along the abscissa indicate that participants' haptic information had greater variability than auditory. The magenta star represents group average performance and shows that with age it goes from being between the worst and best single cue, to being close to best single cue, to being better than best single cue, as predicted by integration. The individual data, however, show that there was a great deal of variability in individual children's performance at the youngest two ages.

Experiment 2

As with Experiment 1, we tested the goodness of fit for each participant and found that the overall mean R^2 for the bimodal congruent condition was 0.87 ± 0.02, for the $+4\Delta$ bimodal incongruent was 0.78 ± 0.02, and for the -4Δ bimodal incongruent was 0.78 ± 0.03, indicating that the chosen Gaussian psychometric function fitted the data well. We carried out a mixed factorial ANOVA with stimulus condition (bimodal congruent and two bimodal incongruent conditions) as within-subjects factor and age group as between-subjects factor on the R^2 values. We found no significant interaction between age and condition ($F(6, 48) = 0.743$, $p = .617$), but a main effect of age ($F(3, 24) = 6.954$, $p = .002$), similar to the findings of Experiment 1. *T*-tests, Bonferroni corrected, post-hoc analyses showed that only the younger group of children differed significantly in their R^2 compared to adults ($p = .002$). Nevertheless, the mean R^2 of the 5–6-year-old children for the bimodal congruent condition was 0.76 $SE \pm 0.08$, i.e. the method of fitting explained the young children's data well.

Before using parametric procedures to test differences in obtained thresholds, we tested whether the data could be approximated by a normal distribution. We performed Kolmogorov–Smirnov tests of normality separately for each age group within each experimental condition. We found that in 11 out of 12 cases the data were normally distributed (i.e. the tests of normality gave non-significant results), and in the remaining case (i.e. the 5–6-year-old children's data in the bimodal congruent condition) the data approximated normality ($p = .043$).

Figure 5 summarizes the results for the bimodal conflict conditions. In Figure 5a the shift in measured and predicted PSEs relative to the 0 conflict condition is plotted against the three levels of haptic-auditory conflict. The slope of the line fitted to these points corresponds to the overall weighting given to haptic as compared with auditory information. A slope of 1 would correspond to complete haptic dominance, slope of -1 to complete auditory dominance (see Figure 5a). All the groups show a slope > 0, i.e. slightly greater weighting for haptic than auditory.

Figure 5b shows the average discrimination threshold obtained for the bimodal congruent and incongruent conditions. To examine the effect of age on the size discrimination threshold for the different conditions, we carried out a mixed model factorial ANOVA with age as a between-subjects factor and condition as within-subjects factor. This analysis revealed a significant main effect of age ($F(3, 24) = 5.460$, $p = .005$), a main effect of condition ($F(2, 48) = 3.588$, $p = .035$), but no significant interaction between age and condition ($F(6, 48) = 0.500$, $p = .805$). Bonferroni post-hoc analyses showed that, overall, adults had significantly lower thresholds than 5–6-year-olds ($p = .005$), but not than either 7–8-year-olds ($p = 1$) or 10–11-year-olds ($p = 1$). The group of 5–6-year-old children had also significantly higher thresholds compared to the 7–8- ($p = .043$), and the 10–11-year-old children ($p = .04$).

Planned *t*-test analyses also indicated that overall the bimodal congruent threshold was significantly lower than both incongruent conditions (-4Δ: $t(27) = -2.323$, $p = .028$; $+4\Delta$: $t(27) = -2.672$, $p = .013$), while no difference in threshold between the two incongruent conditions emerged ($t(27) = 0.390$, $p = .700$). Children and adults behaved similarly and integrated haptic and auditory information by equally weighting the cues. Indeed, the mean weight for each age group was not significantly different from 0.5 (5–6y: $t(5) = 1.061$, $p = .337$; 7–8y: $t(9) = 1.902$, $p = .09$; 10–11y: $t(5) = 0.676$, $p = .529$; adults: $t(5) = 1.290$, $p = .253$).

Discussion

Sensory systems continue to develop during childhood until they reach the mature state. Previous developmental studies have shown that multisensory integration of sensory information involving vision develops quite late (Barutchu *et al.*, 2009; Barutchu *et al.*, 2010; Gori *et al.*, 2008; Innes-Brown *et al.*, 2011; Nardini *et al.*, 2010; Nardini *et al.*, 2008). Children do not reduce the variances of their sensory estimates by integrating information from multiple sources in the same way as adults, until at least 8–10 years of age (Alais & Burr, 2004; Ernst & Banks, 2002; Helbig & Ernst, 2007; Hillis *et al.*, 2002).

However, it remains unknown when multisensory integration of non-visual information such as haptic
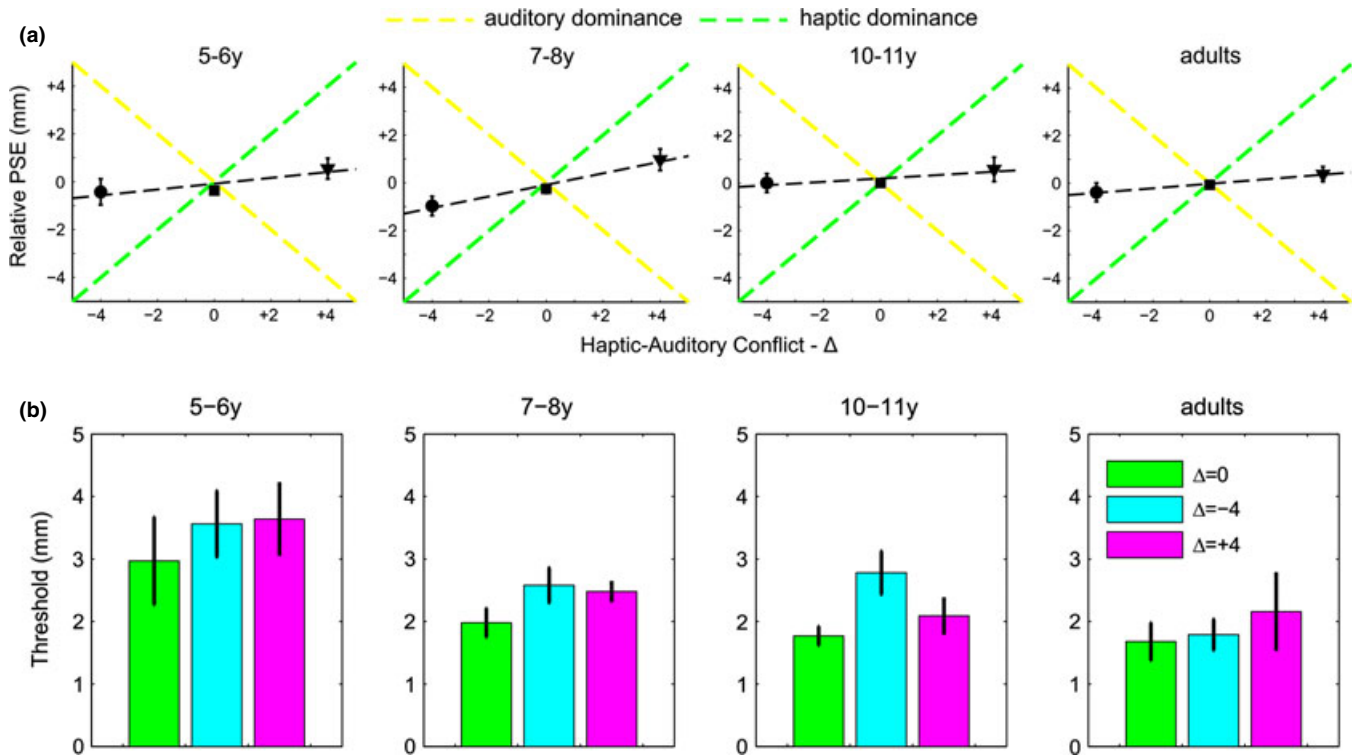
Figure 5 *Predictions and behavior in the conflict condition. (a) The black symbols and dashed line represent the measured PSEs in the conflict condition (on the ordinate) plotted against the three level of haptic-auditory conflict (−4, 0, +4 mm). The yellow dashed line represents the prediction for the auditory dominant model, the green dashed line for the haptic dominant model. Error bars represent standard error of the mean. (b) Mean discrimination thresholds for bimodal congruent (green bar), −4Δ (cyan bar), and +4Δ (magenta bar) conditions as a function of age (different panels). Error bars represent standard error of the mean.*

and auditory develops. Audio-haptic integration for object's size judgment could potentially show either an earlier or a later development than visual-haptic (Gori *et al*., 2008), for reasons we reported in the introduction. Here, we report the findings for two size discrimination experiments to investigate the time course of haptic-auditory integration in the absence of visual information. In Experiment 1 we show that human adults can integrate auditory and haptic cues nearly optimally (i.e. can reduce sensory uncertainty by taking a weighted average of cues) when performing objects' size discrimination. This is consistent with their ability to nearly optimally integrate visual and haptic cues when performing the same kind of task (Gori *et al*., 2008). By 8–10 years, children can integrate visual and haptic cues optimally; however, at the same age children do not show optimal integration of auditory and haptic cues. This result indicates that optimal integration of non-visual cues for objects' size discrimination might occur later in life, at least for the auditory feature under study (i.e. loudness). This conclusion is further supported by the findings of Experiment 2 in which even adults were not

completely able to solve sensory conflict through haptic-auditory integration. To assess whether this holds for other auditory features relevant to size discrimination such as pitch, future studies could use a modified version of our method and set-up. This further investigation would be also important to understand whether the ability to use the pitch information for size discrimination improves or worsens with age.

Our results agree with previous studies (Gori *et al*., 2008; Nardini *et al*., 2010; Nardini *et al*., 2008) in that on average adults, but not children, benefit from combining touch and sound, as indicated by their lower discrimination threshold in this condition when compared to both single-cue conditions. As subjects age, their performance improves (Figure 3); starting worse than the best single cue and ending up better than the best single cue, as predicted by integration. Our results, like those of another recent study (Nardini, Begus & Mareschal, 2012), indicate that some individual participants in all age groups did reduce sensory uncertainty through cue integration (Figure 4). Another result that is consistent with previous reports examining size discrimination in young children

(e.g. Gori *et al.*, 2008) is that the haptic information is weighted more than the other cues when determining both perceived size and discrimination thresholds. With age, this haptic dominance gradually decreases, with adults relying more on the auditory cue when discriminating objects' size.

A further intriguing result is that although 10- to 11-year-old children have lower discrimination thresholds than 7–8-year-olds, their performance on the bimodal condition does not differ from that of the 7- to 8-year-olds. This indicates that the lack of development between these ages is specific to the ability to reduce variability when given both cues. Similar results have been found recently in other studies, and possible explanations of why this happens have been suggested (Barutchu *et al.*, 2009; Barutchu *et al.*, 2010; Nardini *et al.*, 2012). One possibility suggested by Nardini *et al.* (2012) is that during early adolescence the body is growing, and/or sensory systems are developing, more rapidly than at younger ages. From our data, this appears to be especially true for the auditory sensory system. An adult-like ability to use the auditory information to judge object size was reached only at 10 to 11 years, consistent with findings in speaker discrimination (Petrini & Tagliapietra, 2008). The continuing development of unisensory ability may mean that children at these ages are still learning to calibrate their sensory systems and to weight sensory information appropriately (Gori *et al.*, 2010; Gori *et al.*, 2012). Another possibility is that structural and functional reorganization in the brain is at the root of this lack of multisensory development during early adolescence (Paus, 2005; Steinberg, 2005). Physiological and psychophysical studies, examining sight and sound integration in animals and infants, indicate that development of multisensory integration is delayed with respect to that of the two separate senses (Neil, Chee-Ruiter, Scheier, Lewkowicz & Shimojo, 2006; Stein, Labos & Kruger, 1973; Stein, Meredith & Wallace, 1993; Wallace & Stein, 2001).

The results of Experiment 2 show that, on average, both children and adults used similar weighting strategies to solve audio–haptic conflict, resulting in lower thresholds in the absence of conflict. The larger variance, shown by all age groups, when presented with audio–haptic conflict indicates that they were not integrating at all times, but alternated between the cues. One possible explanation behind this lack of integration in all age groups is that participants were aware of the conflict. However, when questioned at the end of the experiment, no participant reported being aware of it. All of the age groups were found to use a similar weighting strategy, equally weighting the haptic and auditory senses. At first glance, this is a very surprising result showing similar behavior in both children and adults. However, the present work differs from previous studies in that it is the first to examine how children and adults would behave when exposed only to multisensory conditions. In other words, the children and adults judging the ball sizes in the three different bimodal conflict conditions were never exposed to either the auditory or haptic information alone. In this situation, which is much closer to real-life situations where experiencing pure single modality is quite rare, all of the age groups similarly weighted (~0.50) the two senses. The similarity between children's and adults' behavior under these circumstances indicates that children do not always weight one modality more than the other, and that their underlying system of cue combination may overall be quite adult-like. We propose that the crucial difference between children and adults is that, unlike adults, individual children do not weight cues according to their own cue reliabilities. This means that while adults and children can show similar mean weights (Figure 5a), and behavior consistent with integration of cues (Figure 5b), children do not show variance reduction given multiple cues (Figure 3). However, some individual children show variance reduction (Figure 4), because they chose appropriate cue weights that reduce variance, perhaps by chance. A similar pattern of results was seen in a recent study of visual–haptic integration for hand localization (Nardini *et al.*, 2012). Individual differences between children who integrate and those who do not could depend on some children learning faster than others to combine the cues, perhaps due to greater exposure to these kinds of sensory contingencies. This possibility could be examined by using learning/training tasks with children.

Intriguingly, many studies have shown that infants possess a variety of multisensory abilities (Lewkowicz, 2000, 2010; Lewkowicz & Ghazanfar, 2009) and young children possess Bayesian-like reasoning abilities (Duffy, Huttenlocher & Crawford, 2006; Gopnik, Glymour, Sobel, Schulz, Kushnir & Danks, 2004; Gopnik & Schulz, 2004; Huttenlocher, Hedges & Duncan, 1991; Huttenlocher, Hedges & Vevea, 2000). For example, 4-year-old children can make Bayesian inferences by combining prior probability information (e.g. almost none of the blocks are blickets) with conditional dependency information (e.g. one of two blocks activates the detector when either alone or with another block) to make causal inferences (only the block activating the detector in both instances is a blicket; Gopnik & Schulz, 2004). Similarly, 5-year-old children can make Bayesian inferences by combining categorical knowledge (e.g. different fish size distributions) with fine-grained information (e.g. fish size) to estimate the sizes of stimuli (matching fish size; Duffy *et al.*, 2006). Thus the ability

to make Bayesian inferences and the capacity to compare information from different senses are available early in development. However, the ability to integrate optimally by choosing the appropriate weights for the different sensory cues develops much later, as shown here and in previous studies (Nardini *et al.*, 2012). This process could stem from the earlier developed Bayesian abilities or all these processes outlined might be quite distinct.

What these findings confirm is that the difference between children and adults lies in the ability to optimally assess and weight the single senses rather than in the ability to integrate them. This appears to be similar for sensory combinations, whether or not they involve vision (Nardini *et al.*, 2012). We do not know whether these developmental changes in weighting strategy reflect consistent changes in underlying multisensory mechanisms. A recent fMRI study demonstrated that in adults haptic and auditory information is integrated at different hierarchical levels in the cortex including superior temporal sulcus (STS; Kassuba *et al.*, 2012). The multisensory process then culminates in the left fusiform gyrus, which seems to be a higher-order region of sensory convergence for object recognition (Kassuba, Klinge, Holig, Menz, Ptito, Roder & Siebner, 2011; Kassuba *et al.*, 2012). In both monkeys and humans, posterior STS is now considered to be a key brain area for multisensory integration (Beauchamp, Lee, Argall & Martin, 2004; Calvert, 2001). In humans, this area responds more to auditory-tactile stimuli than either auditory or tactile in isolation (Beauchamp, Yasar, Frye & Ro, 2008), and more to audio-visual stimuli than either auditory or visual in isolation (Beauchamp *et al.*, 2004). Between 4 and 21 years the posterior portion of the superior temporal gyrus STG develops and undergoes a gradual grey matter loss (Gogtay, Giedd, Lusk, Hayashi, Greenstein, Vaituzis, Nugent, Herman, Clasen, Toga, Rapoport & Thompson, 2004). A loss of grey matter may mean a functional specialization and refinement of this region to achieve adult-like multisensory processing. It may also signify that different multisensory processes develop at different rates during childhood and throughout adolescence until either a common system or different specialized systems are formed. A lack of physiological, psychophysical and neuroimaging studies comparing the time courses for different cross-modal processes from an early age prevent us from knowing whether different task-dependent multisensory processes develop before others in typical individuals. Our study suggests that for the same task, optimal integration for audio-haptic information occurs later than that of visual-haptic (Gori *et al.*, 2008), supporting the existence of separate multisensory mechanisms. However, results from different laboratories cannot be directly compared as they have used different methods and different participants. Future studies may help to elucidate this point, providing a baseline from which to compare the multisensory development of atypical populations and those with sensory deficits.

Acknowledgements

References

Alais, D., & Burr, D. (2004). The ventriloquist effect results from near-optimal bimodal integration. *Current Biology*, **14** (3), 257–262.

Barutchu, A., Crewther, D.P., & Crewther, S.G. (2009). The race that precedes coactivation: development of multisensory facilitation in children. *Developmental Science*, **12** (3), 464–473.

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. *Journal of Experimental Child Psychology*, **105** (1–2), 38–50.

Beauchamp, M.S., Lee, K.E., Argall, B.D., & Martin, A. (2004). Integration of auditory and visual information about objects in superior temporal sulcus. *Neuron*, **41** (5), 809–823.

Beauchamp, M.S., Yasar, N.E., Frye, R.E., & Ro, T. (2008). Touch, sound and vision in human superior temporal sulcus. *NeuroImage*, **41** (3), 1011–1020.

Brainard, D.H. (1997). The Psychophysics Toolbox. *Spatial Vision*, **10** (4), 433–436.

Bresciani, J.P., & Ernst, M.O. (2007). Signal reliability modulates auditory–tactile integration for event counting. *NeuroReport*, **18** (11), 1157–1161.

Calvert, G.A. (2001). Crossmodal processing in the human brain: insights from functional neuroimaging studies. *Cerebral Cortex*, **11** (12), 1110–1123.

Demorest, S.M. (1992). Information integration theory: an approach to the study of cognitive development in music. *Journal of Research in Music Education*, **40** (2), 126–138.

Demorest, S.M., & Serlin, R.C. (1997). The integration of pitch and rhythm in musical judgment: testing age-related trends in novice listeners. *Journal of Research in Music Education*, **45** (1), 67–79.

Duffy, S., Huttenlocher, J., & Crawford, L.E. (2006). Children use categories to maximize accuracy in estimation. *Developmental Science*, **9** (6), 597–603.

Ernst, M.O., & Banks, M.S. (2002). Humans integrate visual and haptic information in a statistically optimal fashion. *Nature*, **415** (6870), 429–433.

Gogtay, N., Giedd, J.N., Lusk, L., Hayashi, K.M., Greenstein, D., Vaituzis, A.C., Nugent, T.F. 3rd, Herman, D.H., Clasen, L.S., Toga, A.W., Rapoport, J.L., & Thompson, P.M. (2004). Dynamic mapping of human cortical development during childhood through early adulthood. *Proceedings of the National Academy of Sciences, USA*, **101** (21), 8174–8179.

Gopnik, A., Glymour, C., Sobel, D.M., Schulz, L.E., Kushnir, T., & Danks, D. (2004). A theory of causal learning in children: causal maps and Bayes nets. *Psychological Review*, **111** (1), 3–32.

Gopnik, A., & Schulz, L. (2004). Mechanisms of theory formation in young children. *Trends in Cognitive Sciences*, **8** (8), 371–377.

Gori, M., Del Viva, M., Sandini, G., & Burr, D.C. (2008). Young children do not integrate visual and haptic form information. *Current Biology*, **18** (9), 694–698.

Gori, M., Sandini, G., Martinoli, C., & Burr, D. (2010). Poor haptic orientation discrimination in nonsighted children may reflect disruption of cross-sensory calibration. *Current Biology*, **20** (3), 223–225.

Gori, M., Tinelli, F., Sandini, G., Cioni, G., & Burr, D. (2012). Impaired visual size-discrimination in children with movement disorders. *Neuropsychologia*, **50** (8), 1838–1843.

Grassi, M. (2005). Do we hear size or sound? Balls dropped on plates. *Perception & Psychophysics*, **67** (2), 274–284.

Helbig, H.B., & Ernst, M.O. (2007). Optimal integration of shape information from vision and touch. *Experimental Brain Research*, **179** (4), 595–606.

Hillis, J.M., Ernst, M.O., Banks, M.S., & Landy, M.S. (2002). Combining sensory information: mandatory fusion within, but not between, senses. *Science*, **298** (5598), 1627–1630.

Huttenlocher, J., Hedges, L.V., & Duncan, S. (1991). Categories and particulars: prototype effects in estimating spatial location. *Psychological Review*, **98** (3), 352–376.

Huttenlocher, J., Hedges, L.V., & Vevea, J.L. (2000). Why do categories affect stimulus judgment? *Journal of Experimental Psychology: General*, **129** (2), 220–241.

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. *Developmental Science*, **14** (5), 1089–1099.

Kassuba, T., Klinge, C., Holig, C., Menz, M.M., Ptito, M., Roder, B., & Siebner, H.R. (2011). The left fusiform gyrus hosts trisensory representations of manipulable objects. *NeuroImage*, **56** (3), 1566–1577.

Kassuba, T., Menz, M.M., Roder, B., & Siebner, H.R. (2012). Multisensory interactions between auditory and haptic object recognition. *Cerebral Cortex*, **23** (5), 1097–1107.

Lewkowicz, D.J. (2000). Infants' perception of the audible, visible, and bimodal attributes of multimodal syllables. *Child Development*, **71** (5), 1241–1257.

Lewkowicz, D.J. (2010). Infant perception of audio-visual speech synchrony. *Developmental Psychology*, **46** (1), 66–77.

Lewkowicz, D.J., & Ghazanfar, A.A. (2009). The emergence of multisensory systems through perceptual narrowing. *Trends in Cognitive Sciences*, **13** (11), 470–478.

Nardini, M., Bedford, R., & Mareschal, D. (2010). Fusion of visual cues is not mandatory in children. *Proceedings of the National Academy of Sciences, USA*, **107** (39), 17041–17046.

Nardini, M., Begus, K., & Mareschal, D. (2012). Multisensory uncertainty reduction for hand localization in children and adults. *Journal of Experimental Psychology: Human Perception and Performance*, **39** (3), 773–787.

Nardini, M., Jones, P., Bedford, R., & Braddick, O. (2008). Development of cue integration in human navigation. *Current Biology*, **18** (9), 689–693.

Nava, E., & Pavani, F. (2013). Changes in sensory dominance during childhood: converging evidence from the colavita effect and the sound-induced flash illusion. *Child Development*, **84** (2), 604–616.

Neil, P.A., Chee-Ruiter, C., Scheier, C., Lewkowicz, D.J., & Shimojo, S. (2006). Development of multisensory spatial integration and perception in humans. *Developmental Science*, **9** (5), 454–464.

Paus, T. (2005). Mapping brain maturation and cognitive development during adolescence. *Trends in Cognitive Sciences*, **9** (2), 60–68.

Pelli, D.G. (1997). The VideoToolbox software for visual psychophysics: transforming numbers into movies. *Spatial Vision*, **10** (4), 437–442.

Petrini, K., & Tagliapietra, S. (2008). Cognitive maturation and the use of pitch and rate information in making similarity judgments of a single talker. *Journal of Speech, Language and Hearing Research*, **51** (2), 485–501.

Sanabria, D., Soto-Faraco, S., & Spence, C. (2005). Assessing the effect of visual and tactile distractors on the perception of auditory apparent motion. *Experimental Brain Research*, **166** (3–4), 548–558.

Soto-Faraco, S., & Deco, G. (2009). Multisensory contributions to the perception of vibrotactile events. *Behavioural Brain Research*, **196** (2), 145–154.

Stein, B.E., Labos, E., & Kruger, L. (1973). Sequence of changes in properties of neurons of superior colliculus of the kitten during maturation. *Journal of Neurophysiology*, **36** (4), 667–679.

Stein, B.E., Meredith, M.A., & Wallace, M.T. (1993). The visually responsive neuron and beyond: multisensory integration in cat and monkey. *Progress in Brain Research*, **95**, 79–90.

Steinberg, L. (2005). Cognitve and affective development in adolescence. *Trends in Cognitve Sciences*, **9** (2), 69–74.

Tajadura-Jimenez, A., Valjamae, A., Toshima, I., Kimura, T., Tsakiris, M., & Kitagawa, N. (2012). Action sounds recalibrate perceived tactile distance. *Current Biology*, **22** (13), R516–R517.

Wallace, M.T., & Stein, B.E. (2001). Sensory and multisensory responses in the newborn monkey superior colliculus. *Journal of Neuroscience*, **21** (22), 8886–8894.