What Box: a task for assessing language lateralisation in young children

Nicholas A Badcock[1], Rachael Spooner[2], Jessica Hofmann[3], Atlanta Flitton[3], Scott Elliott[3], Lisa Kurylowicz[3,4], Louise M Lavrencic[3], Heather M Payne[5,6], Georgina K Holt[7], Anneka Holden[7], Owen F Churches[7], Mark J Kohler[2], and Hannah AD Keage[3].

[1] ARC Centre for Excellence in Cognition and its Disorders, Department of Cognitive Science, Macquarie University, North Ryde, New South Wales, Australia

[2] Cognitive Neuroscience Laboratory, School of Psychology, Social Work and Social Policy, University of South Australia, South Australia, Australia

[3] Cognitive Ageing and Impairment Neurosciences Laboratory, School of Psychology, Social Work and Social Policy, University of South Australia, South Australia, Australia.

[4] Department of General Paediatrics, University Children's Hospital Bonn, Bonn, Germany

[5] Deafness, Cognition, and Language Research Centre, University College London, United Kingdom

[6] Institute of Cognitive Neuroscience, University College London, United Kingdom

[7] Oxford Study of Children's Communication Impairments, Department of Experimental Psychology, University of Oxford, United Kingdom

[7] Brain and Cognition Laboratory, Flinders University, South Australia, Australia

Corresponding Author:

Nicholas A Badcock

16 University Avenue, North Ryde, New South Wales, Australia

Email address: nicholas.badcock@mq.edu.au

Acknowledgements

This research was supported by a programme grant from the Wellcome Trust (ref 082498/Z/07/Z) and ARC Centre of Excellence Grant [CE110001021]. We would also like to thank Oxford Study of Children's Communication Impairments (OSCCI) for their contributions to the development of the task and Margriet A. Groen, Eva Guiterrez-Sigut, Metten Somers, and our anonymous reviewers for their feedback on the manuscript.

Abstract

The assessment of active language lateralisation in infants and toddlers is challenging. It requires an imaging tool that is unintimidating, quick to setup, and robust to movement, in addition to an engaging and cognitively simple language processing task. Functional Transcranial Doppler Ultrasound (fTCD) offers a suitable technique and here we report on a suitable method to elicit active language production in young children. The 34-second 'What Box' trial presents an animated face 'searching' for an object. The face 'finds' a box that opens to reveal a to-be-labelled object. In a sample of 95 children (1 to 5 years-of-age), 81% completed the task – 32% with ≥ 10 trials. The task was validated ($\rho = 0.4$) against the gold standard Word Generation task in a group of older adults (n = 65, 60 to 85 years-of-age), though was less likely to categorise lateralisation as left or right, indicative of greater measurement variability. Existing methods for active language production have been used with 2-year-old children while passive listening has been conducted with sleeping 6-month-olds. This is the first active method to be successfully employed with infants through to pre-schoolers, forming a useful tool for populations in which complex instructions are problematic.

Keywords

Language, lateralisation, functional Transcranial Doppler Ultrasound, infants, toddlers

Introduction

The specialisation of cognitive capacities to the left and right cerebral hemispheres is referred to as the lateralisation of cognitive function and, in most people, the left hemisphere is specialised (or dominant) for language processing whilst the right is specialised for visuo-spatial processing. Early in development, there is evidence of this specialisation for language reception (Dehaene-Lambertz, 2000) but the lateralisation of language production has been harder to determine. Here we report a method for examining language reception and production that is suitable for use with young children.

Owing to the inherent difficulty for children below the age of 5 to stay still – a significant problem for functional Magnetic Resonance Imaging (fMRI) – researchers have favoured functional Transcranial Doppler Ultrasound (fTCD) for investigating language lateralisation in this age group. FTCD is used to measure the blood flow velocity in the left and right cerebral arteries, most commonly, the middle cerebral arteries (Aaslid, Markwalder, & Nornes, 1982; Newell & Aaslid, 1992); faster event-related velocities in a given hemisphere are indicative of cerebral lateralisation for that event (i.e., language production). The gold standard task for assessing language lateralisation using fTCD is Word Generation - visually cued word generation (Knecht et al., 1996). The task is reliable (Knecht, Deppe, Ringelstein, et al., 1998; Stroobant & Vingerhoets, 2001), and has been validated against Wada (Knake et al., 2003; Knecht, Deppe, Ebner, et al., 1998) and fMRI (Knecht, Deppe, Ebner, et al., 1998; Somers, Neggers, Kahn, & Sommer, 2011). However, whilst Word Generation works well for adults, silent word production to letters (i.e., requiring letter-sound knowledge) and long periods of inanimate relaxation are not suitable for children. Alternatives have been developed.

Child-friendly fTCD tasks include Picture Description (Haag et al., 2010; Lohmann, Drager, Muller-Ehrenberg, Deppe, & Knecht, 2005), Animation Description (Bishop, Watt, & Papadatou-Pastou, 2009), and Story Listening (Stroobant, Van Boxstael, & Vingerhoets, 2011). These have been used with children as young as two-years-of-age but continue to rely on sustained periods of rest and attention. Picture Description and Story Listening include approximately 30 seconds of production or listening followed by 30 seconds of rest. The animation description task is more child-friendly with 12 seconds of animation following by 10 seconds of production and 8 seconds of rest; however, our pilot work determined that this task was not suitable to maintain 18-month-olds' interest. In addition, the reliance on overt production is difficult for children below the age of four.

Covert language has been used to successfully activate the cerebral structures involved in overt production (Bookheimer et al., 1998) and the strength of lateralisation is similar for covert and overt production (Gutierrez-Sigut, Payne, & MacSweeney, 2015). Taking advantage of this, Wilke et al. (2005) developed tasks that induce the automatic covert production of predictable words that are replaced within sentences by a tone. For example, "A frog lived under a flower. One day a girl picked the [tone]." Observers automatically fill-in the missing word as evidenced by increased activity in areas usually associated with overt production. This activity is enhanced by the presentation of a picture of the missing word. This covert production task has been successfully completed by children as young as six-years-of-age using fMRI, producing plausible lateralisation indices (Lidzba, Schwilling, Grodd, Krägeloh-Mann, & Wilke, 2011). Using fTCD, this task has been compared with Word Generation in adults but lateralisation was weaker and less reliable than Word Generation (Badcock, Nye, & Bishop, 2012a). However, participants were not given instructions and the paradigm did not explicitly encourage labelling.

The What Box Task

The ‘What Box’ task follows from this literature as a procedure to elicit covert or overt language production in young children. Here we build upon a previous report of the task (Kohler et al., 2015), providing a detailed methodology for the presentation and administration of the task as well as updated processing and analysis techniques for use with fTCD in young children and older adults. In the adults, What Box was also compared with the gold standard Word Generation task.

Materials and Methods

Participants

Children

Ninety-five children between 1- and 5-years-of-age (M = 39.46 months, SD = 15, min = 12, max = 67), born between 35- and 42-weeks’ gestation, were tested. Forty-nine (52%) were male. Children were included if English was the primary language, they had no known visual or auditory impairments, learning problems, developmental delays or syndromes affecting cognitive development (e.g., autism or Down syndrome), they were not currently taking medication known to affect cardiovascular blood vessel function or neurocognitive performance (such as a stimulant or psychotropic drug) or suffering from any acute illness, such as a cold.

Ninety percent of the sample were Caucasian with socioeconomic status (M = 1009.2, SD = 47.9) similar to the national mean (M = 1000, SD = 100: Australian Bureau of Statistics Index of Relative Socio-economic Advantage/Disadvantage 2011 national census data). Hand preference was determined by planned observation of the use of age-appropriate objects, based on methods used in children from 6-months of age (Michel, Ovrut, & Harkins, 1985): 76 (80%) were right-handed, 12 (12.63%) were left-handed, and 7 (7.37%) did not demonstrate a dominant hand.

Older adults

Sixty-seven adults with a mean age of 68.94 years (SD = 6, min = 60, max = 85) also participated. Twenty-eight (42%) were male. All were right-handed based on the Flinders Handedness Survey (Nicholls, Thomas, Loetscher, & Grimshaw, 2013).

Procedure

What Box

The What Box task includes an animation of a face 'searching' for an object. The animation is created with a series of still-frame images and accompanying sounds (see Figure 1 for a schematic diagram of a trial including timing). The key steps are:

1. A blank background is presented
2. The face 'moves' down and then up the screen

3. A box appears then opens followed by a spoken "Look!"
4. The box is then replaced by an object and a spoken "What's this?", and the object's verbal label is presented after a delay, to allow for verbal labelling
5. A face with hands covering its mouth appears with the spoken "Shh"

What Box was administered to children and adults in the same manner with the exception that a different set of stimuli was used for each group and the task was discontinued after 20 trials with a correct response for adults. There were 51 stimuli, chosen from http://websites.psychology.uwa.edu.au/school/MRCDatabase/uwa_mrc.htm (for a list see Supplementary Materials). There was a minimum of 25 trials in adults, and 37 were required for two individuals to achieve 20 correct labels. Additional stimulus and presentation details are provided as supplementary materials.

Insert Figure 1 about here

Word Generation

Adult participants also completed the Word Generation task: based on Knecht et al. (1996). There were 24, 60 sec trials corresponding with the letters of the alphabet, excluding 'x' and 'z'. Each trial consisted of six periods (note: words in inverted commas were displayed on the screen and acted as instructions): 1. A blank normalisation period (15 sec), 2. 'Clear Mind' (5 secs), 3. a single, randomly selected letter was presented on the screen (2.5 sec), 4. silent word generation of words beginning with the presented letter (12.5 secs), 5. 'Say' (5 secs), and 6. 'Relax' (15 secs). Brief auditory tones were presented at the start of the clear mind, say, and relax periods.

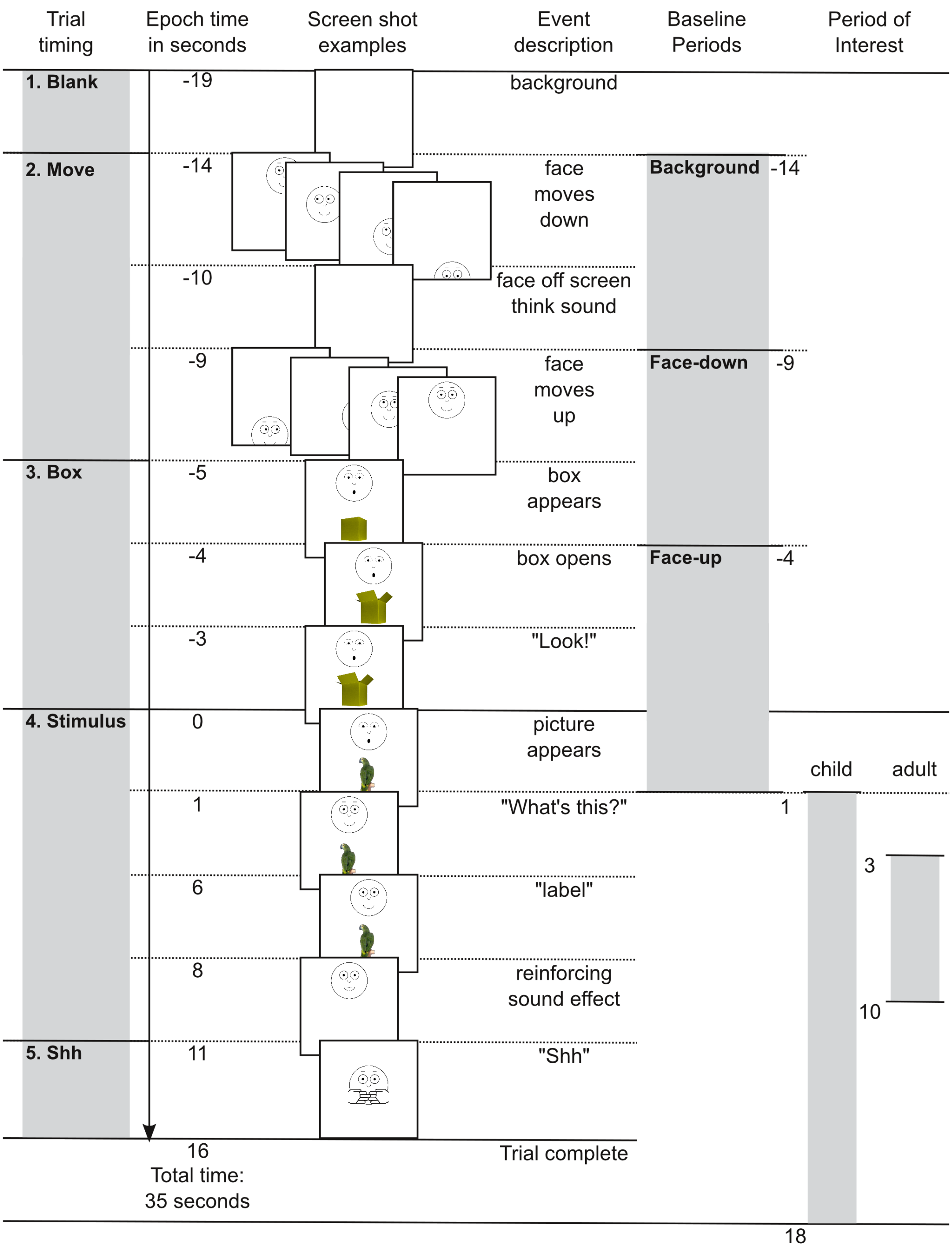


Figure 1. A schematic diagram of a What Box trial. Includes trial timing, event descriptions, and baseline and period of interest timings for children and adult data processing. Please note, the timing of the baseline periods accounts for the delay due to neurovascular-coupling.

Functional Transcranial Doppler Ultrasonography

Blood flow velocity was measured using a Doppler ultrasonography device (Doppler-Box™, DWL Elektronische Systeme, Singen, Germany) with a Diamon® headset or elastic headbands to hold 2-MHz transducer probes over the left and right temporal skull windows to insonate the middle cerebral arteries. Seated participants viewed the Psychtoolbox (Brainard, 1997; Pelli, 1997) MATLAB (R2011b Mathworks, Natick, MA, USA) presentation on a 22-inch Dell P2210 monitor (50 cm viewing distance). Event markers were inserted via parallel port (http://apps.usd.edu/coglab/psyc770/IO32.html). FTCD data collection was conducted after standardised test administration as part of ongoing research at the University of South Australia, Cognitive Neuroscience Laboratory (for preliminary findings see Keage et al., 2015; Kohler et al., 2015).

Data processing

The fTCD data were processed using DOPOSCCI (Badcock, Holt, Holden, & Bishop, 2012) version 3.0, a MATLAB-based summary-suite for fTCD data (see https://github.com/nicalbee/dopStep). DOPOSCCI implements the processing described by Deppe et al. (Deppe, Knecht, Henningsen, & Ringelstein, 1997; Deppe, Knecht, Lohmann, & Ringelstein, 2004). Here we extend upon these methods to maximise epoch retention and reliability. The data were trimmed to exclude irrelevant recording before the first and after the final epoch. Heart cycle artefacts were removed (see Deppe et al., 1997, 2004) and smoothed using MATLAB's 'linspace' function across steps.

Epochs were excluded manually if the participant was observed to be disengaged from the task, conducted gross movements, or was talking during the baseline period. The median number of manually excluded epochs was 2 (IQR = 6, min = 0, max = 20).

To maximise data retention, spurious extreme values beyond -3 or 4 standard deviations from the mean[1], affecting less than 5% of the data, were adjusted using 'linspace' between values 1.5 secs either side of the extreme value[2]. The data were epoched from baseline onset (see below and Figure 1) to 18 secs relative to event markers, and normalised to a mean of 100 within each epoch (i.e., not overall), correcting for left and right probe angle differences (see Deppe et al., 2004). Baseline correction was conducted, subtracting the mean of data within this period (see below) from all other data points.

Epochs with extreme values were excluded: values beyond ± 50% of the mean or with a left-minus-right difference greater than 8%, affecting more than 1% of the data within the epoch. Regarding the left-minus-right difference (or activation separation), the literature reports group fTCD laterality indices in the magnitude of 3 to 5% change (e.g., Mean + 3xSD = 7.33, Badcock, Nye, & Bishop,

---

[1] These values are based on a personal communication from Dorothy Bishop, found to be effective for the correction of movement-related signal disruption from children's fTCD data.

[2] For future reference, the descriptives for the affected data sets were: <-3 SD, n = 37 (48%), Median = 10 % data affected, IQR = 35, min = 0.016, max = 98; > 4 SD, n = 2 (2%), Median = 0.43, IQR = 0.06, min = 0.36, max = 5.

2012b), therefore, separations greater than this are likely due to movement artefacts. See supplementary materials for justification of the 8% criterion.

Baseline selection

As this is a new paradigm, and bearing in mind the 5 to 7 sec delay due to the timing of neurovascular coupling (Malonek et al., 1997; Rosengarten, Osthaus, & Kaps, 2002) – i.e., the stimulus-related neurophysiological response will be delayed, therefore the timing of these baselines is delayed - three baseline periods were tested in children to determine the most suitable using split-half reliability as an index of quality (see supplementary materials). The three baseline periods are displayed in Figure 1. The 'face-up' period was selected as the most reliable. This was -4 to 1 secs relative to event onset, including activation to the presentation of the face moving up the screen.

Laterality Index (LI) calculation and categorisation

Laterality Indices (LIs) were calculated as the average left minus right signal over a 2-sec period surrounding the peak left-right difference within the period of interest: 5 to 18 secs in children (see below for adults). Positive LI values indicate left lateralisation, negative indicate right.

To determine whether the LI was significantly different to zero, a one-sample t-test was applied to the LI values for the group. Split-half reliability was calculated based upon LIs calculated for the odd and even numbered epochs, adjusted to equate the number of odd and even epochs used.

LIs were also categorised as left, right, or neither based on the overlap of 95% confidence intervals with zero (i.e., an LI was considered left if the lower interval was greater than zero). Categorisation comparisons were conducted using Chi-squared and McNemar (i.e., repeated measures Chi-squared) tests. These comparisons tested whether the ratios differed: 1) within children dependent upon the selection of epochs (< 10 versus ≥ 10), 2) between children and adults for the What Box task, and 3) within adults for between task comparisons. Regarding 3) McNemar tests require binomial categories, therefore two of the three categories were collapsed for each comparison: left versus not (neither + right), right versus not, and neither vs not.

Older adult data processing

What Box

Alternate timings were used for the older adults' What Box data: epoch -14 to 10 secs, baseline -14 to -9, and period of interest 3 to 10. As evidenced by the physiological response (see Figure 2, Panel B), the adults adhered to the instruction better than the children, requiring alternate timing. The baseline period was earlier, corresponding to 10 seconds after the 'Shh' instruction (see Figure 1 trial schematic and timing). The period of interest was earlier and shorter, longer periods picked up a second component in some individuals resulting in changes from typical to atypical lateralisation and poorer internal reliability.

For the adults, epoch exclusion by activation separation was based on individually calculated cut-offs. The distribution of separations was smaller for adults than children – average median = 3.01 (IQR = 1.59, Min = 0.97, Max = 9.81) – indicative of cleaner recordings. The median activation separation plus eight times the interquartile range was most reliable method of screening epochs for activation separation, increasing the split-half reliability from ρ = .65 [95% CI: .45 .78] without screening to ρ = .71 [.52 .84]. Spearman's rank order correlations were used to reduce the impact of extreme values.

Word Generation

The Word Generation data were processed as described above with timings based on previous research (Keage et al., 2015; Knecht et al., 1996; Knecht, Deppe, Ringelstein, et al., 1998); epoch -15 to 25 secs, baseline -15 to -5, and period of interest 5 to 15. Individually calculated cut-offs were used for activation separation epoch exclusion, five times the inter-quartile range (for reference, the average median activation separation was 3.80, IQR = 2.89, Min = 1.16, Max = 13.06). This cut-off increased the split-half reliability from ρ = .77 [95% CI: .63 .86] without screening to ρ = .82 [.69 .89].

What Box and Word Generation comparison

Data from participants with 10 or more accepted epochs for both What Box and Word Generation were included in the data analysis (there were two exclusions). Validity was calculated by disattenuating (Schumacker & Muchinsky, 1996; Spearman, 1904) the correlations between the LIs for the two tasks.

Results

What Box

For the What Box task, the group-averaged change in blood flow velocity, for the left and right middle cerebral arteries, relative to object presentation is displayed in Figure 2 (children and adults in Panels A & B respectively).

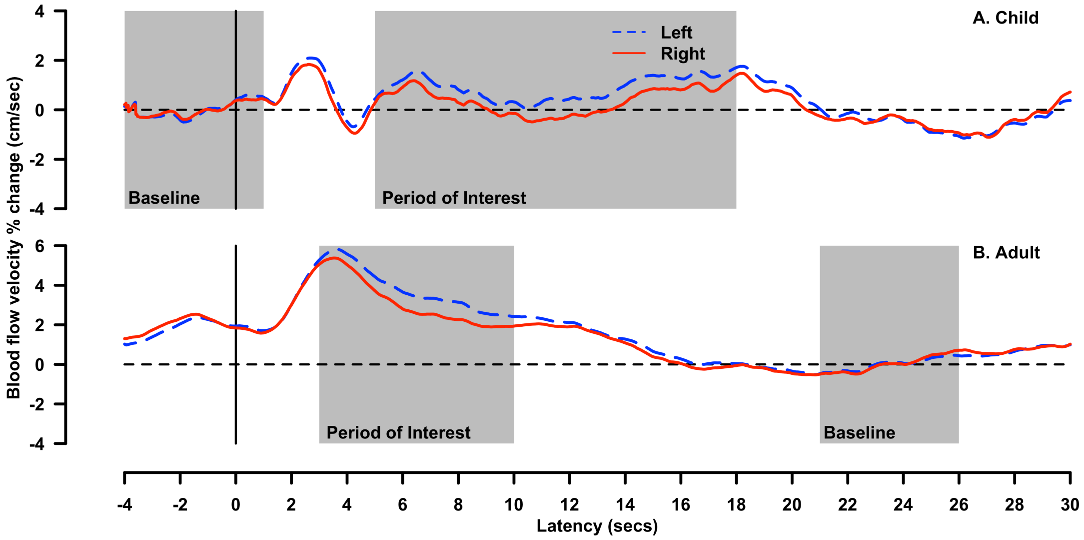


Figure 2. Group-averaged change in blood flow velocity relative to object presentation (Latency = 0 seconds) for the left (broken blue line) and right (solid red line) as a function of time (in seconds). Panel A displays the infant data (n = 77, 81% of the total sample) that were calculated using a -4 to 1 sec baseline period (first grey panel). Panel B displays the adult data (n = 66, 99% of total sample) that were calculated with a 21 to 26 sec baseline (equivalent to -14 to -9 but adjusted for visualisation here to maintain the same x-axis). The periods of interest (-5 to 18 secs for infants and 3 to 10 for adults) are also displayed for reference. Please note the y-axis range is greater in panel B.

For the children, there are three features to note. The first feature is an early (around 3 secs), non-lateralised peak that likely reflects a rapid, attention-related response to the object presentation. The second feature includes two, left-lateralised peaks (around 6.5 and 16 secs) that likely reflect a labelling response to the object and a receptive or repetition response to the verbally presented label. These peaks are included in the period of interest. The third feature is convergence of the left and right velocities: evident at 22 seconds. This has implications for the selection of the baseline period. The continuation of task-related activity into the 'Blank' phase of the next trial (see Figure 1) has an impact on the task reliability, dependent upon the timing of the baseline period; i.e., this continuation produces poorer reliability for the -14 to -9 baseline compared to -4 to 1 that does not have this continuation, which aids in the justification of its use.

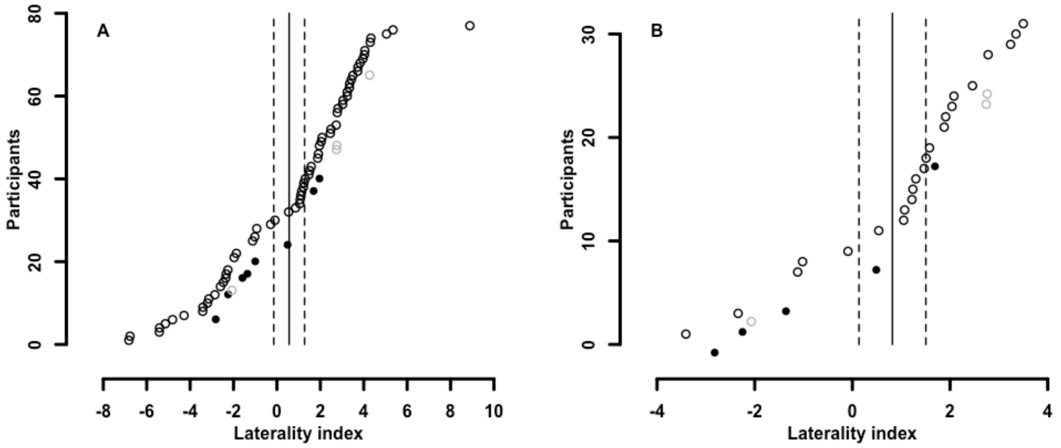


Figure 3. The distribution of laterality indices for A) participants with 1 or more accepted epochs (n = 77, 81% of the total sample) and B) participants with 10 or more accepted epochs (n = 31, 33% of the total sample). Sample mean (solid vertical line) and 95% confidence intervals (dashed vertical lines) are also displayed. Filled and greyed symbols represent left-handed and undetermined-handed individuals respectively, offset vertically for visualisation.

There were 1 or more acceptable epochs for 77 participants (81% of the total sample): median = 7, IQR = 10, min = 1, max = 32. The distribution of all laterality indices (LIs) is displayed in Figure 3, panel A, and the categorisations (i.e., left, neither, right) are presented in Table 1. The number of accepted epochs was correlated with age such that older children had more accepted epochs, Spearman's $\rho = 0.38$ [0.18 0.59], $p < .01$.

Table 1

Lateralisation categorisation numbers (percentages) by Task, Sample, and Epoch selection.

Sample	Task	Epochs	n	Left	Neither	Right
Child	What Box	All available	77	29 (37.66)	27 (35.06)	21 (27.27)
		Less than 10	46	14 (30.43)	17 (36.96)	15 (32.61)
		10 or more	31	15 (48.39)	10 (32.26)	6 (19.35)
Adult	What Box	10 or more	65	32 (49.23)	22 (33.85)	11 (16.92)
	Word Gen	10 or more	65	40 (61.54)	16 (24.62)	9 (13.85)

The minimum number of acceptable epochs for LI calculations varies in the literature from 8 (Gutierrez-Sigut et al., 2015) to 12 (Groen, Whitehouse, Badcock, & Bishop, 2011): here we used 10. Based on this criterion, the distribution of LIs for participants with 10 or more epochs is displayed in Figure 3, panel B (n = 31, 33% of the total sample). The number of accepted epochs (median = 14, IQR = 6, min = 10, max = 32) was not significantly related to age, Spearman's $\rho = 0.06$ [-0.32 0.45], $p = 0.75$. The mean LI was 0.82 ($SD = 1.95$, min = -3.41, max = 3.5, 95%CI = 0.68), which is statistically different to zero $t(30) = 2.35$, $p = 0.026$; and represents a medium effect size, Cohen's d = 0.42. On average, the group was left-dominant for language processing. Laterality categorisations are presented in Table 1 and were not significantly affected by epoch selection (i.e., < 10 versus ≥ 10); $\chi^2(2, N = 77) = 2.89$, $p = 0.24$, Cramer's V = 0.14 (small to medium effect). The split-half reliability is 0.64 [0.37 0.81], $t(29) = 4.47$, $p < .001$. In addition, the What Box laterality categorisations did not differ significantly between children and adults: χ^2 (2, N = 96) = 0.09, *p* = 0.96, Cramer's V = 0.02.

What Box versus Word Generation

There were 65 older adult participants with 10 or more epochs for both the What Box and Word Generation tasks (What Box: median = 24, IQR = 3, min = 15, max = 27; Word Generation: median = 22, IQR = 3, min = 11, max = 24). The mean LI for both tasks indicated left lateralisation overall: What Box = 0.95 (SD = 2.36, latency = 6.39, latency SD = 2.37, *t*(64) = 3.25, *p* < .01, d = 0.40), Word Generation = 1.57 (SD =2.47, latency = 9.31, latency SD = 2.75, *t*(64) = 5.13, *p* < .001, d = 0.64). The internal reliability for both tasks was high (What Box, ρ = 0.71, Word Generation, ρ = 0.82) and the disattenuated correlation between the two tasks was $\rho = 0.40$, indicating medium correspondence. A scatter plot of the LIs for the two tasks is presented in Figure 4.

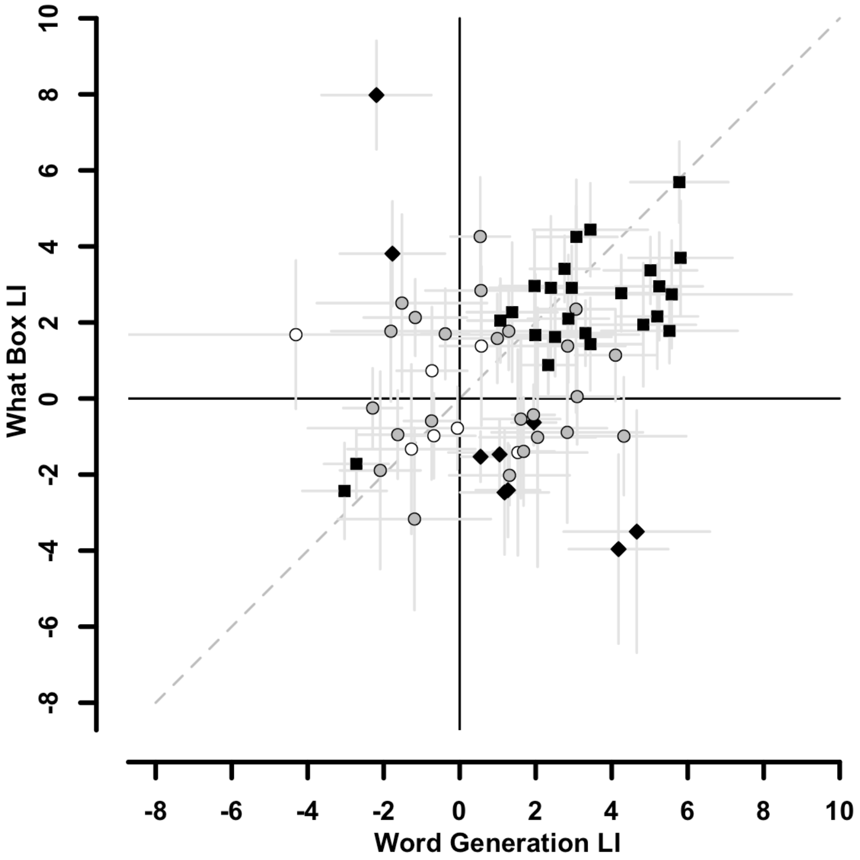


Figure 4. Scatter plot of the laterality indices (LIs) for the What Box and Word Generation tasks (n = 65, 97% of the total sample). A diagonal line is included for reference to consistent LI mapping between the task. Lateralisation categorisation (based on the 95% confidence interval criterion used for Table 1) is indicated by symbols-colour combinations: black symbols indicate statistically reliable categorisation in both tasks (square = left or right on both, and diamond = switched between tasks - i.e., left to right or right to left), grey diamonds indicate a neither categorisation in one task and left or right in the other, and open circles indicate neither categorisation in both tasks. The 95% confidence intervals are displayed for each individual (light grey).

The laterality categorisations for the two tasks are presented in Table 1. McNemar tests indicated that, the ‘neither’ categorisation was 1.3 to 1.6 times more likely in What Box versus Word Generation: categorisation did not differ for left versus not (neither or right); χ^2 (1, N = 65) = 0.49, p = 0.48, OR = 0.61; but did for right versus not (left or neither), χ^2(1, N = 65) = 30.73, p < .001, OR = 1.27; and for neither versus not (right or left), χ^2(1, N = 65) = 11.46, p <.001, OR = 1.57.

Discussion

Here we report the methods and statistical characteristics of a child-friendly task for the assessment of language lateralisation using fTCD. The task presentation involves a face ‘looking’ for something, finding a box, the box opening, and an object appearing. Observers are prompted with “What’s this?” and the label of the object, cueing overt and/or covert language production. This was successfully employed with young children aged between 1 and 5 years as well as older adults (60 to 85). Laterality indices (LIs) showed a broad distribution, with the group averages indicative of left-lateralisation. In addition, the older adults also completed the gold standard fTCD assessment for language lateralisation, Word Generation (Knecht et al., 1996; Knecht, Deppe, Ebner, et al., 1998). The LIs for both tasks were correlated ($\rho = 0.40$), indicative of medium evidence for the What Box task invoking language processing. In addition, the task was less likely to categorise individuals as left or right dominant for language processing (50% correspondence with Word Generation), suggesting it to be a more variable paradigm overall.

The rates of lateralisation categorisation for Word Generation are different relative to previous work with right-handed populations (92.5% left and 7.5% right in Knecht et al., 2000; vs 61.5% left, 24.5% neither, and 14% right, see Table 1). Whilst the paradigms are the same and the current sample speaks English rather than German in the earlier work, the major difference between the populations is the age: between 17 and 50 in Knecht et al., average of 26; versus between 60 and 85, average of 69 in the current study. This may be the critical factor as language lateralisation is reported to reduce with aging (Keage et al., 2015; Matteis et al., 1998). The is consistent with increased rates of non-left lateralisation reported here and may speak to developmental changes late, as well as earlier in development – language lateralisation may stabilise around the age of reading instruction (Groen, Whitehouse, Badcock, & Bishop, 2012) but this is a fertile space of enquiry. These developmental changes may also contribute to the weaker lateralisation indices observed.

The work adds to the methods available for assessing lateralisation using fTCD in children, including Picture Naming (Haag et al., 2010; Lohmann et al., 2005), Story Listening (Stroobant et al., 2011), and Animation Description (Bishop et al., 2009). Relative to the existing techniques, the internal reliability for the What Box – r = 0.64 [0.37 0.81] – was lower than Animation Description (r = .89 to .90 in 4-year-olds; Bishop, Holt, Whitehouse, & Groen, 2014; Bishop et al., 2009) and lower but comparable to Picture Naming depending upon the study (r = .88, Lohmann et al., 2005; Intra-class correlation = .66, Stroobant et al., 2011). It should be noted that the average number of accepted epochs was lower for What Box than other tasks and the internal reliability was higher when more suitable epochs were available (n = 12, r = 0.69 [0.36, 0.87]; n = 14, r = 0.76 [0.36, 0.92]; see Supplementary Materials). The fact that the current sample included younger children than other studies (previously down to 2 years of age), does not entirely account for this discrepancy as the adult sample also demonstrated lower reliability compared with Word Generation.

We employed two novel approaches to data exploration and retention using DOPOSCCI (Badcock, Holt, et al., 2012). This included activation correction and activation separation epoch screening (for specific details see the method and Supplementary Materials). These are little-explored forms of data cleaning and screening that maximised reliability but would benefit from replication and refinement.

We tested three baseline periods to establish the best processing methods for the What Box task: 1. -14 to -9 secs (time relative to stimulus), the presentation of a background image; 2. -9 to -4 secs, presentation of the animated face moving down the computer monitor; and 3. -4 to 1 secs, presentation of the animated face moving up the computer monitor. Relative to the end ‘Shh’ instruction of the previous trial, these periods were 0, 5, and 10 secs respectively. Examination of the reliability for each baseline as a function of the number of acceptable epochs indicated that the latest period was most consistent (-4 to 1, 10 secs after following the end of the previous trial). With reference to Figure 2, this is not surprising; the left-minus-right difference has normalised (i.e., no difference) by 10 seconds after the end of the previous trial. This is in line with neurovascular coupling estimates (Rosengarten et al., 2002).

Future work may benefit from increasing the duration of the face-animation stages of the paradigm.

Limitations

Despite the What Box producing left-lateralisation at the group level for children and adults, the index was relatively weak (i.e., lower lateralisation indices) and only moderately correlated with Word Generation is adults. The neural substrates underpinning Word Generation have been demonstrated (e.g., Deppe et al., 2000), whereas this is unknown for What Box, and it is possible that alternate activation may underpin the differences. In relation to the patterns of activation (i.e., Figure 2), it is clear that qualitative differences exist between the children and adults. We speculate that this relates to the adult's consistent approach to the task (i.e., a single alerting and production peak) whereas children appear to be alerted and then cued to production – perhaps involving greater executive processes – and then production extends over a longer time-course (i.e., two peaks and up-to 8 secs longer than adults for left-right channel convergence). In addition, differences may be due to the low volume of production required in the task. Recently Payne and colleagues (2015) demonstrated that reduced rates of production are associated with weaker lateralisation. This pattern of behaviour likely accounts for the weaker lateralisation observed for the What Box task. Increasing the number of to-be-labelled objects per trial may increase the lateralisation index as well as the correspondence between What Box and Word Generation.

Based on the adult literature, a minimum of 8 (Gutierrez-Sigut et al., 2015) to 12 (Groen et al., 2011) epochs are suggested for LI calculation. In the current paper, we used 10. To date, there has been no empirical test to determine the optimal number of epochs but the data presented here indicate that higher internal reliability is associated with a greater number of epochs, therefore, we recommend a minimum of 10 epochs for methods such as What Box. We suggest multiple testing sessions to achieve these numbers as well as tailoring stimuli to the interests of each individual participant if on-task behaviour is poor. In addition, the criterion against which to judge the validity of a task is currently against the gold standard Word Generation task. If lateralisation were clearly predictive of some behaviour, this would provide a better criterion against which to judge the quality of a task, and in turn, the cost of minimal epochs. Currently, this remains elusive (Bishop, 2013).

Future applications

Although What Box was designed for typically developing infants and toddlers, we also demonstrated its use with older adults. The task is simple and may be conducted without verbal instructions. This provides a rare paradigm that can be applied across a broad age-range to map the development of lateralisation. Given the flexibility of the task, it will be useful in populations with atypical development such as dyslexia, specific language impairment, and Autism; where research has previously used the Word Generation in adults (dyslexia, Illingworth & Bishop, 2009; specific language impairment and autism, Whitehouse & Bishop, 2008). In addition, the simplicity of What Box makes it useful for working with populations where memory for and adherence to the rules associated with Word Generation limit its application; including, intellectual impairment (e.g., Down syndrome, Bowler, Cufflin, & Kiernan, 1985), cognitive decline such as aging (Keage et al., 2015), dementia (Matteis et al., 1998), and brain damage (Bragoni et al., 2000). TCD per se has been applied successfully in a wide range of populations (for systematic reviews see Bakker et al., 2014 in children, Keage et al., 2012 in aging and dementia), therefore the combination of fTCD and What Box provides a useful tool.

Conclusion

We report detailed methodology and data processing for the assessment of language lateralisation in young children that can also be used with adults. The method, the 'What Box' task, was successfully employed in children aged between 1 and 5 years using functional Transcranial Doppler Ultrasounds (fTCD) and showed medium correspondence with Word Generation collected with older adults.

References

Aaslid, R., Markwalder, T. M., & Nornes, H. (1982). Noninvasive transcranial Doppler ultrasound recording of flow velocity in basal cerebral arteries. *Journal of Neurosurgery*, *57*(6), 769–774. https://doi.org/10.3171/jns.1982.57.6.0769

Badcock, N. A., Holt, G., Holden, A., & Bishop, D. V. M. (2012). dopOSCCI: a functional Transcranial Doppler Ultrasonography summary suite for the assessment of cerebral lateralization of cognitive function. *Journal of Neuroscience Methods*, *204*(2), 383–388. https://doi.org/10.1016/j.jneumeth.2011.11.018

Badcock, N. A., Nye, A., & Bishop, D. V. M. (2012a). Using functional transcranial Doppler ultrasonography to assess language lateralisation: Influence of task and difficulty level. *Laterality*, *17*(6), 694–710. https://doi.org/10.1080/1357650X.2011.615128

Badcock, N. A., Nye, A., & Bishop, D. V. M. (2012b). Using functional transcranial Doppler ultrasonography to assess language lateralisation: Influence of task and difficulty level. *Laterality*, *17*(6), 694–710. https://doi.org/10.1080/1357650X.2011.615128

Bakker, M. J., Hofmann, J., Churches, O. F., Badcock, N. A., Kohler, M., & Keage, H. A. (2014). Cerebrovascular function and cognition in childhood: a systematic review of transcranial doppler studies. *BMC Neurology*, *14*(43). https://doi.org/10.1186/1471-2377-14-43

Bishop, D. V. M. (2013). Cerebral Asymmetry and Language Development: Cause, Correlate, or Consequence? *Science*, *340*(6138), 1230531. https://doi.org/10.1126/science.1230531

Bishop, D. V. M., Holt, G., Whitehouse, A. J. O., & Groen, M. (2014). No population bias to left-hemisphere language in 4-year-olds with language impairment. *PeerJ*, *2*, e507. https://doi.org/10.7717/peerj.507

Bishop, D. V. M., Watt, H., & Papadatou-Pastou, M. (2009). An efficient and reliable method for measuring cerebral lateralization during speech with functional transcranial Doppler ultrasound. *Neuropsychologia*, *47*(2), 587–590. https://doi.org/10.1016/j.neuropsychologia.2008.09.013

Bookheimer, S. Y., Zeffiro, T. A., Blaxton, T. A., Gaillard, W. D., Malow, B., & Theodore, W. H. (1998). Regional cerebral blood flow during auditory responsive naming: evidence for cross-modality neural activation. *Neuroreport*, *9*(10), 2409–2413.

Bowler, D. M., Cufflin, J., & Kiernan, C. (1985). Dichotic listening of verbal and non verbal material by Down's syndrome children and children of normal intelligence. *Cortex; a Journal Devoted to the Study of the Nervous System and Behavior*, *21*(4), 637–644.

Bragoni, M., Caltagirone, C., Troisi, E., Matteis, M., Vernieri, F., & Silvestrini, M. (2000). Correlation of cerebral hemodynamic changes during mental activity and recovery after stroke. *Neurology*, *55*(1), 35–40.

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

Dehaene-Lambertz, G. (2000). Cerebral specialization for speech and non-speech stimuli in infants. *Journal of Cognitive Neuroscience*, *12*(3), 449–460.

Deppe, M., Knecht, S., Henningsen, H., & Ringelstein, E. B. (1997). AVERAGE: a Windows® program for automated analysis of event related cerebral blood flow. *Journal of Neuroscience Methods*, *75*(2), 147–154. https://doi.org/16/S0165-0270(97)00067-8

Deppe, M., Knecht, S., Lohmann, H., & Ringelstein, E. B. (2004). A method for the automated assessment of temporal characteristics of functional hemispheric lateralization by transcranial Doppler sonography. *Journal of Neuroimaging:*

Official Journal of the American Society of Neuroimaging, *14*(3), 226–230. https://doi.org/10.1177/1051228404264936

Deppe, M., Knecht, S., Papke, K., Lohmann, H., Fleischer, H., Heindel, W., … Henningsen, H. (2000). Assessment of Hemispheric Language Lateralization: A Comparison between fMRI and fTCD. *Journal of Cerebral Blood Flow & Metabolism*, *20*(2), 263–268. https://doi.org/10.1097/00004647-200002000-00006

Groen, M. A., Whitehouse, A. J. O., Badcock, N. A., & Bishop, D. V. M. (2011). Where were those rabbits? A new paradigm to determine cerebral lateralisation of visuospatial memory function in children. *Neuropsychologia*, *49*(12), 3265–3271. https://doi.org/16/j.neuropsychologia.2011.07.031

Groen, M. A., Whitehouse, A. J. O., Badcock, N. A., & Bishop, D. V. M. (2012). Does cerebral lateralization develop? A study using functional transcranial Doppler ultrasound assessing lateralization for language production and visuospatial memory. *Brain and Behavior*, *2*(3), 256–269. https://doi.org/10.1002/brb3.56

Gutierrez-Sigut, E., Payne, H., & MacSweeney, M. (2015). Investigating language lateralization during phonological and semantic fluency tasks using functional transcranial Doppler sonography. *Laterality: Asymmetries of Body, Brain and Cognition*, *20*(1), 49–68. https://doi.org/10.1080/1357650X.2014.914950

Haag, A., Moeller, N., Knake, S., Hermsen, A., Oertel, W. H., Rosenow, F., & Hamer, H. M. (2010). Language lateralization in children using functional transcranial Doppler sonography. *Developmental Medicine and Child Neurology*, *52*(4), 331–336. https://doi.org/10.1111/j.1469-8749.2009.03362.x

Illingworth, S., & Bishop, D. V. M. (2009). Atypical cerebral lateralisation in adults with compensated developmental dyslexia demonstrated using functional transcranial Doppler ultrasound. *Brain and Language*, *111*(1), 61–65. https://doi.org/10.1016/j.bandl.2009.05.002

Keage, H. A. D., Churches, O. F., Kohler, M., Pomeroy, D., Luppino, R., Bartolo, M. L., & Elliott, S. (2012). Cerebrovascular Function in Aging and Dementia: A Systematic Review of Transcranial Doppler Studies. *Dementia and Geriatric Cognitive Disorders EXTRA*, *2*(1), 258–270. https://doi.org/10.1159/000339234

Keage, H. A. D., Kurylowicz, L., Lavrencic, L. M., Churches, O. F., Flitton, A., Hofmann, J., … Badcock, N. A. (2015). Cerebrovascular function associated with fluid, not crystallized, abilities in older adults: A transcranial Doppler study. *Psychology and Aging*, *30*(3), 613–623. https://doi.org/10.1037/pag0000026

Knake, S., Haag, A., Hamer, H. M., Dittmer, C., Bien, S., Oertel, W. H., & Rosenow, F. (2003). Language lateralization in patients with temporal lobe epilepsy: a comparison of functional transcranial Doppler sonography and the Wada test. *NeuroImage*, *19*(3), 1228–1232.

Knecht, S., Deppe, M., Dräger, B., Bobe, L., Lohmann, H., Ringelstein, E.-B., & Henningsen, H. (2000). Language lateralization in healthy right-handers. *Brain*, *123*(1), 74–81. https://doi.org/10.1093/brain/123.1.74

Knecht, S., Deppe, M., Ebner, A., Henningsen, H., Huber, T., Jokeit, H., & Ringelstein, E. B. (1998). Noninvasive Determination of Language Lateralization by Functional Transcranial Doppler Sonography: A Comparison With the Wada Test. *Stroke*, *29*(1), 82–86. https://doi.org/10.1161/01.STR.29.1.82

Knecht, S., Deppe, M., Ringelstein, E. B., Wirtz, M., Lohmann, H., Dräger, B., … Henningsen, H. (1998). Reproducibility of Functional Transcranial Doppler Sonography in Determining Hemispheric Language Lateralization. *Stroke*, *29*(6), 1155–1159.

Knecht, S., Henningsen, H., Deppe, M., Huber, T., Ebner, A., & Ringelstein, E. B. (1996). Successive activation of both cerebral hemispheres during cued word generation. *Neuroreport*, *7*(3), 820–824. https://doi.org/10.1097/00001756-199602290-00033

Kohler, M., Keage, H. A. D., Spooner, R., Flitton, A., Hofmann, J., Churches, O. F., … Badcock, N. A. (2015). Variability in lateralised blood flow response to language is associated with language development in children aged 1–5 years. *Brain and Language*, *145–146*, 34–41. https://doi.org/10.1016/j.bandl.2015.04.004

Lidzba, K., Schwilling, E., Grodd, W., Krägeloh-Mann, I., & Wilke, M. (2011). Language comprehension vs. language production: Age effects on fMRI activation. *Brain and Language*, *119*(1), 6–15. https://doi.org/10.1016/j.bandl.2011.02.003

Lohmann, H., Drager, B., Muller-Ehrenberg, S., Deppe, M., & Knecht, S. (2005). Language lateralization in young children assessed by functional transcranial Doppler sonography. *NeuroImage*, *24*(3), 780–790. https://doi.org/10.1016/j.neuroimage.2004.08.053

Malonek, D., Dirnagl, U., Lindauer, U., Yamada, K., Kanno, I., & Grinvald, A. (1997). Vascular imprints of neuronal activity: Relationships between the dynamics of cortical blood flow, oxygenation, and volume changes following sensory stimulation. *Proceedings of the National Academy of Sciences*, *94*(26), 14826–14831.

Matteis, M., Silvestrini, M., Troisi, E., Bragoni, M., Vernieri, F., & Caltagirone, C. (1998). Cerebral hemodynamic patterns during stimuli tasks in multi-infarct and Alzheimer types of dementia. *Acta Neurologica Scandinavica*, *97*(6), 374–380.

Michel, G. F., Ovrut, M. R., & Harkins, D. A. (1985). Hand-use preference for reaching and object manipulation in 6- through 13-month-old infants. *Genetic, Social, and General Psychology Monographs*, *111*(4), 407–427.

Newell, D. W., & Aaslid, R. (1992). Transcranial Doppler: clinical and experimental uses. *Cerebrovascular and Brain Metabolism Reviews*, *4*(2), 122–143.

Nicholls, M. E. R., Thomas, N. A., Loetscher, T., & Grimshaw, G. M. (2013). The Flinders Handedness survey (FLANDERS): A brief measure of skilled hand preference. *Cortex*, *49*(10), 2914–2926. https://doi.org/10.1016/j.cortex.2013.02.002

Payne, H., Gutierrez-Sigut, E., Subik, J., Woll, B., & MacSweeney, M. (2015). Stimulus rate increases lateralisation in linguistic and non-linguistic tasks measured by functional transcranial Doppler sonography. *Neuropsychologia*, *72*, 59–69. https://doi.org/10.1016/j.neuropsychologia.2015.04.019

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

Rosengarten, B., Osthaus, S., & Kaps, M. (2002). Overshoot and undershoot: control system analysis of haemodynamics in a functional transcranial Doppler test. *Cerebrovascular Diseases*, *14*(3–4), 148–152. https://doi.org/10.1159/000065672

Schumacker, R., & Muchinsky, P. (1996). Disattenuating correlation coefficients. *Rasch Measurement Transactions*, *10*(1), 479.

Somers, M., Neggers, S. F., Kahn, R. S., & Sommer, I. E. (2011). The measurement of language lateralization with functional transcranial Doppler and functional MRI: a critical evaluation. *Frontiers in Human Neuroscience*, *5*, 31. https://doi.org/10.3389/fnhum.2011.00031

Spearman, C. (1904). The proof and measurement of association between two things. *American Journal of Psychology*, *15*(1), 72–101.

Stroobant, N., Van Boxstael, J., & Vingerhoets, G. (2011). Language lateralization in children: A functional transcranial Doppler reliability study. *Journal of Neurolinguistics*, *24*(1), 14–24. https://doi.org/10.1016/j.jneuroling.2010.07.003

Stroobant, N., & Vingerhoets, G. (2001). Test-retest reliability of functional transcranial Doppler ultrasonography. *Ultrasound in Medicine & Biology*, *27*(4), 509–514.

Whitehouse, A. J. O., & Bishop, D. V. M. (2008). Cerebral dominance for language function in adults with specific language impairment or autism. *Brain*, *131*(12), 3193–3200. https://doi.org/10.1093/brain/awn266

Wilke, M., Lidzba, K., Staudt, M., Buchenau, K., Grodd, W., & Krägeloh-Mann, I. (2005). Comprehensive language mapping in children, using functional magnetic resonance imaging: what's missing counts. *Neuroreport*, *16*(9), 915–919.

What Box: a task for assessing language lateralisation in young children – Supplementary Materials

Procedure

Stimuli

The visual stimuli include backgrounds, faces, boxes, and objects. The backgrounds were coloured photographs including houses, rooms (e.g., kitchen, bedroom), and natural scenes (e.g., gardens, landscapes). Images were blurred and mirrored: blurring reduced the presence of attention-capturing, high-contrast features and mirroring (along the vertical centre) controlled for any bias in the lateralisation of visual attention between hemifields. Some of these images did contain nameable objects, however, the degradation of the images and context meant that we did not observe evidence of overt or covert labelling to these backgrounds. The faces were blue in colour, included two eyes and eyebrows, a nose, and a mouth. Black pupils were set in white eyes and pupil position was varied to adjust gaze direction (centred, up, down, left, or right). Eye-shape was altered from full circles to horizontal crescents to indicate surprise (see ‘box appears’ event in Figure 1). The mouth shape included a u-shaped line smile; a vertically oriented, black oval to indicate surprise; and a horizontal u-shaped, white crescent ‘smile’ as reinforcement for monitoring the display – not accuracy. Images of open and shut cardboard boxes were presented in 14 different colours (aqua, light and dark blue, brown, light and dark green, orange, pink, purple, red, rust, turquoise, white, and yellow). There were 33 different images of objects (e.g., biscuit, bottle, and animals, for a full list see Supplementary Table 1) selected as items commonly known by 18-month-old children (from the Oxford Study of Children's Communication Impairments databases).

The auditory stimuli included: spoken labels for each of the objects, "Look!" recorded with rising intonation to capture attention, "What's this?", and a series of sound-effects: 13 action files (e.g., spring, cork pop, or whistle), 3 to indicate 'thinking' (e.g., Hmm), and 26 celebratory sounds used for reinforcement (e.g., crowd cheers, "yahoo", "yay", and laughing). The spoken words were recorded in a female British accent.

Supplementary Table 1

Stimulus list for infants and adults. Items are reported in columns by presentation order.

Infants		Adults		
baby	hat	clever	armour	sapphire
ball	horse	bread	scorpion	honey
banana	house	diamond	anchor	telephone
bath	light	scissors	key	camel
bed	milk	clover	bed	clock
bird	plane	owl	elephant	medal
biscuit	shoe	cabbage	mosquito	brick
book	sock	cake	ring	ambulance
bunny	teddy	skunk	pyramid	apple
bus	train	saxophone	ants	
cake	tv	bib	matches	
car	window	caravan	bomb	
cat		radio	limousine	
chair		atom	chair	
cow		scroll	mallet	
cup		cigar	armadillo	
dog		pencil	eggs	
door		dog	pineapple	
duck		plaster	chalk	
fish		shirt	jelly	
foot		cart	lute	

Trial timing

Each trial lasted 35 seconds. The timings will be described in 5 periods relative to the animation, including the duration of the period and the timing relative to the presentation of the object at time 0 (for a diagram see Figure 1).

1. Blank (5 sec, -19 to -14): A randomly selected background was presented for 5 sec then remained as the background until the object appeared (i.e., time 0).
2. Move (9 sec, -14 to -5): The face stimulus was presented at locations simulating movements down then up the screen. The location changed at 1 sec intervals and was accompanied by a randomly selected action sound. There were four down and four up vertical locations randomly varied to be within the four vertical quarters of the screen. The horizontal positions were varied to left or right of centre within a corridor 20% of the screen width. This corridor was used to avoid any bias in the lateralisation of visual attention. The position of the eyes also varied randomly at each position (i.e., looking left, right, up, and down) except for the top position of the screen when they were straight ahead (i.e., looking at the participant). Following the downward movement, the face 'moved' off the bottom of the screen for 1 sec, accompanied by a 'thinking' sound. Following the upward movement, the face always finished horizontally centred in the top quarter of the screen.

3. Box (5 sec, -5 to 0): A box was presented and opened with an action sound at each step, 1 sec between each step, and the face looked down and surprised. The “Look!” cue was then presented to direct attention to the screen, 3 sec.
4. Stimulus (11 sec, 0 to 11): The object was presented on a black background (the face remained in the top central position looking surprised and straight ahead at the participant), for 1 sec during which an event marker was sent for data analysis. The “What’s this?” cue was then played. After 5 secs (allowing for word generation/production) the object auditory label was played. After 2 sec a smiling face, with a reinforcing sound effect, was presented and remained on screen for 3 secs. The objects were presented in alphabetical order.
5. Shh (5 sec, 11 to 16): A larger face with hands over its mouth was then presented for 5 sec accompanied by a ‘Shh’ sound.

Familiarisation with the Transcranial Doppler Ultrasound (TCD) headset

Familiarisation included demonstrating the equipment on one of the two or three researchers present as well as the parent, and allowing the child to play with and decorate the headset with stickers. If necessary a teddy bear ‘helper’ was fitted with the headset and read the book ‘I can hear my brain’ with the child (see Appendix 1). Children sat on a chair or on their caregiver’s lap (younger, < 3, and non-compliant children), watching a favourite television programme, while the headset or headband was then fixed in place and probes attached. The headset was a better fit for older children, while the headbands were more suitable for younger children or those with asymmetric heads. Upon the accurate detection of the MCA (confirmed by bifurcation checking when possible) on the left and right side, the probes were fixed in place.

We recommend the following steps for optimal insonation:

1. Brush hair backwards and out of the way
2. Ask the child to yawn whilst looking at the temple – this can give a good indication of the ‘dint’ or thinnest part of the skull.
3. Begin searching for the MCA at the following location: making reference to the outer canthus of the eye, move posteriorly to the hairline, to above the zygomatic arch.
4. Position the probe to be facing slightly upwards and forwards, towards the back of the contralateral eye.
5. Using small steps, move the probe around an imaginary clock-face to find the best point of insonation.
6. Increase software gain and reference the M-Mode spectrograph to determine optimal depth and position.
7. Increase the depth of the pulse to find the MCA bifurcation (indicated by bi-directional flow in the spectrograph) and reduce depth until the M1 section of the MCA is reached. This is where the cleanest signal should be found.

Gel selection

Adhesive conductive gel (Tensive® by Parker) or Echoson® Ultrasonographic Gel (Sonogel Vertriebs GmbH) were used in the current experiment. The choice of gel depended on the age and compliance of the child:

the adhesive gel being used with younger and less compliant children as it can be placed on the temporal window without running which is more convenient for setup.

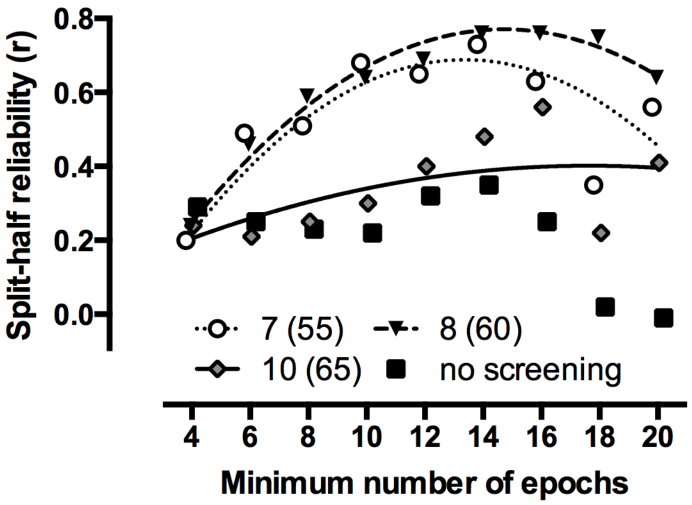
1.1.1 Testing session

Following standardised test administration, each participant was familiarised and fitted with the TCD headset. The task was introduced as a game with the aim of naming objects in a box that a face finds. The instructions were delivered in developmentally appropriate language including: i) the requirement to wait until something comes out of the box and ii) to label the object that comes out of the box. The first trial was used as practice to ensure the participant understood the requirements of the task. If necessary, the participant's attention was re-directed to the task throughout testing, and any gross motor movements or diversion from the task was recorded for manual epoch exclusion.

Data processing

Activation separation (Left-minus-right channel difference)

As a form of epoch screening, epochs with an extreme activation separation – or left-minus-right difference – were excluded. The chosen cut-off of 8% was based on the 60th percentile (8.12) of the median left-minus-right difference of the sample (average median difference was 6.6%, interquartile range = 11.39%). This was set by examining the split-half reliability (correlation between laterality indices calculated for the odd and even epochs) as a function of the minimum number of epochs included in the calculation at separations of 7% and 10% (55th and 65th percentiles of the average median left-minus-right differences respectively), and with all available epochs (i.e., no screening). The number of available epochs varied between individuals and depended upon the activation separation screening. Without screening, the reliability was poor. At an 8% cut-off, the reliability was strongest (see Supplementary Figure 1). Second-order polynomial equations ($y = B_0 + B_1x + B_2x^2$) differentiated the 8% and 7% separation from the 10%; $F(6,18) = 4.46$, $p < .001$ (no screening was not included in the analysis); with R^2 values of .99, .84, and .49 respectively (see Supplementary Table 2 for parameter statistics). The same curve adequately fitted the 8% and 7% separations, $F(3,12) = 1.02$, $p < .42$. We used the more inclusive cut-off: 8%. These summaries at based on a -4 to 1 baseline period. Please note, median values were used so as to avoid the influence of epochs with extreme values.



Supplementary Figure 1. Split-half reliability (Pearson product moment r values) for four levels of left-minus-right activation separation (7, 8, 10, and no screening, numbers reflecting the 55th, 60th, and 65th percentiles of the median difference of the sample respectively) as a function of the minimum number of epochs included in the calculation. The best fitting quadratic regression lines are displayed for separations of 7 (dotted line), 8 (dashed line), and 10 (solid line).

Supplementary Table 2

Second-order polynomial parameter statistics for activation separation cut-offs as a function of the minimum number of epochs included in the calculation.

Separation % (%ile)	Parameter		
	B0	B1	B2
7 (55)	-0.26 [-0.6 0.07]	0.14 [0.06 0.22]	-0.005 [-0.009 -0.002]
8 (60)	-0.21 [-0.31 -0.12]	0.13 [0.11 0.16]	-0.005 [-0.006 -0.004]
10 (65)	0.07 [-0.28 0.43]	0.04 [-0.04 0.11]	-0.001 [-0.004 0.002]

Baseline selection

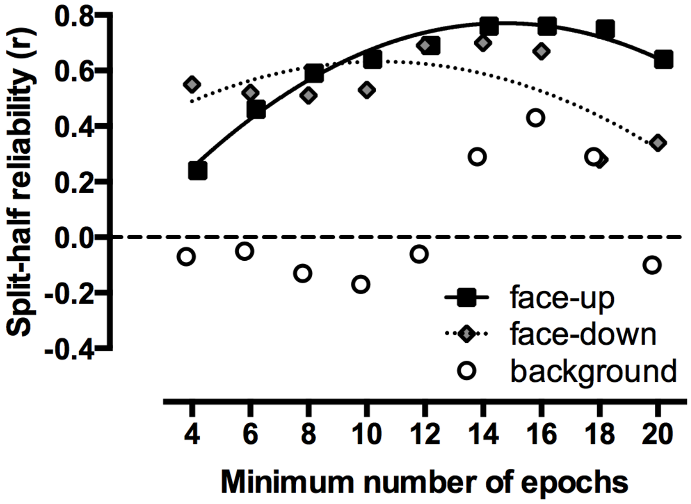
As this is a new paradigm, we tested three baseline periods to determine the most suitable, using split-half reliability as an index of quality, bearing in mind the 5 to 7 sec delay due to the timing of neurovascular coupling (Malonek et al., 1997; Rosengarten, Osthaus, & Kaps, 2002). That is, the neurophysiological response related to the presentation of a particular stimulus or period will be delayed, therefore the timing of these baselines is delayed. The three baseline periods were:

1) ‘background’, -14 to -9 secs, including activity to the presentation of the background, commencing 10 secs after the onset of ‘Shh’;

2) ‘face-down’, -9 to -4 secs, including activity to the presentation of the face moving down the screen, commencing 10 secs after the onset of the background; and

3) ‘face-up’, -4 to 1 secs, including activity to the presentation of the face moving up the screen, commencing 10 secs after the onset of the face.

Split-half reliabilities were calculated for the three baseline periods for a range of epochs: 2 to 10 for odd and even epoch halves (i.e., at least 4 to 20 acceptable epochs in total). These reliabilities are displayed in Supplementary Figure 2 (for a complete set of the summary statistics for these divisions including sample size, LI estimates, and reliability confidence intervals, see Supplementary Table 3). Second-order quadratic equations ($y = B_0 + B_1x + B_2x^2$, see Supplementary Table 4 for best fitting parameter statistics, conducted with GraphPad Prism 6.0f) were fitted to the reliabilities to evaluate the relative suitability of the baseline periods. The reliabilities were higher and more consistent for the ‘face-up’ baseline: best fitting values differentiated the ‘face-up’ and ‘face-down’ baselines; $F(3, 12) = 7.42$, $p < 0.01$. Therefore the -4 to 1 baseline period was deemed most suitable for this paradigm. In addition to reliability, this baseline retained a greater number of epochs across participants; likely due to shorter epoch duration, and fewer epochs rejected for artefacts (see Data processing above for epoch rejection criteria).



Supplementary Figure 2. Split-half reliability (Pearson product moment r values) for three baseline periods (background = -14 to -9, face-down = -9 to -4, and face-up = -4 to 1) as a function of the minimum number of epochs included in the calculation. The best fitting quadratic regression lines are displayed for face-down (dotted line) and face-up (solid line) data.

Supplementary Table 3

Split-half reliability (Pearson product moment r values), 95% confidence intervals (CI), and laterality index (LI) descriptive statistics for three baseline periods (background = -14 to -9, face-down = -9 to -4, and face-up = -4 to 1) as a function of the minimum number of epochs included in the calculation. The descriptive statistics include: n = the number of participants included in the calculations, and LI values: mean and standard deviation (SD), and median and inter-quartile range (IQR).

Baseline Period	Min Epochs	n	LI mean (SD)	LI median (IQR)	r [95% CI]
Face-up	4	51	0.83 (2.24)	1.24 (3.54)	0.24 [-0.03, 0.49]
[-4 to 1]	6	40	0.92 (2.02)	1.23 (3.12)	0.46 [0.17, 0.67]
	8	33	0.71 (1.93)	1.24 (3.04)	0.59 [0.31, 0.78]
	10	29	0.87 (1.97)	1.3 (2.17)	0.64 [0.36, 0.82]
	12	20	0.97 (2.02)	1.4 (2.19)	0.69 [0.36, 0.87]
	14	13	0.8 (2.13)	1.3 (1.39)	0.76 [0.36, 0.92]

Baseline Period	Min Epochs	n	LI mean (SD)	LI median (IQR)	r [95% CI]
	16	13	0.8 (2.13)	1.3 (1.39)	0.76 [0.36, 0.92]
	18	8	1.53 (1.78)	1.78 (1.75)	0.75 [0.1, 0.95]
	20	5	2.15 (1.33)	2.46 (2.19)	0.64 [-0.56, 0.97]
Face-down	4	45	0.7 (2.16)	1.32 (3.58)	0.55 [0.31, 0.73]
[-9 to -4]	6	36	0.68 (1.91)	1.23 (3.2)	0.52 [0.23, 0.73]
	8	33	0.63 (1.92)	1.14 (3.24)	0.51 [0.2, 0.72]
	10	26	0.82 (1.81)	1.38 (2.89)	0.53 [0.18, 0.76]
	12	18	0.24 (1.82)	0.52 (3.29)	0.69 [0.33, 0.88]
	14	11	0.16 (2.02)	0.3 (3.34)	0.7 [0.17, 0.91]
	16	10	0.44 (1.89)	0.98 (2.84)	0.67 [0.07, 0.91]
	18	7	0.98 (1.53)	1.65 (2.2)	0.28 [-0.6, 0.85]
	20	4	0.84 (1.37)	1 (1.8)	0.34 [-0.92, 0.98]
Background	4	43	0.03 (1.81)	0.1 (3.06)	-0.07 [-0.36, 0.24]
[-14 to -9]	6	35	0.11 (1.64)	0.28 (2.82)	-0.05 [-0.37, 0.29]
	8	32	0.05 (1.61)	0.19 (2.61)	-0.13 [-0.45, 0.23]
	10	23	-0.12 (1.44)	-0.32 (2.44)	-0.17 [-0.54, 0.26]
	12	18	0.07 (1.4)	0.38 (2.32)	-0.06 [-0.51, 0.42]
	14	11	0.25 (1.52)	0.98 (2.75)	0.29 [-0.38, 0.76]
	16	10	0.44 (1.46)	1.02 (2.74)	0.43 [-0.28, 0.83]
	18	8	0.45 (1.42)	1.02 (1.96)	0.29 [-0.52, 0.83]
	20	5	0.65 (1.47)	1.07 (2.41)	-0.1 [-0.9, 0.86]

Supplementary Table 4

Second-order polynomial best fitting parameters, 95% confidence intervals, and R^2 values for reliability coefficients calculated for incremental numbers of epochs for two baseline periods.

Baseline	Parameters			
	B_0	B_1	B_2	R^2
-9 to -4	0.26 [-0.28 0.79]	0.07 [-0.03 0.17]	-0.003 [-0.007 0]	0.6
-4 to 1	-0.21 [-0.31 -0.12]	0.13 [0.11 0.16]	-0.005 [-0.006 -0.004]	0.99

FTCD data clipping: recording error

Eight of the recordings were affected by an incorrect software setting that set an upper-limit on the recorded velocity: blood-flow velocities above 133 cm/secs were saved as 133 (automatically detected as > 2% of the signal being equal to the maximum value, see DOPOSCCI ‘dopClipCheck’ function for more information). In order to determine whether this significantly affected the LI calculations, this limit was artificially imposed on the processing of the all other data sets. The percentage of artificially clipped data ranged from 0.01 to 36% with a mean of 11.77%. The LI calculations were not affected; mean difference = -0.19 (SD = 1.79), $t(138) = -0.38$, $p = 0.70$; and showed a strong correspondence, $r = 0.82$, $p < .001$. Therefore, the restricted data were included in the full analysis. These summaries in this paragraph are based on a -4 to 1 baseline period.

References

Malonek, D., Dirnagl, U., Lindauer, U., Yamada, K., Kanno, I., & Grinvald, A. (1997). Vascular imprints of neuronal activity: Relationships between the dynamics of cortical blood flow, oxygenation, and volume changes following sensory stimulation. *Proceedings of the National Academy of Sciences*, *94*(26), 14826–14831.

Rosengarten, B., Osthaus, S., & Kaps, M. (2002). Overshoot and undershoot: control system analysis of haemodynamics in a functional transcranial Doppler test. *Cerebrovascular Diseases*, *14*(3–4), 148–152. https://doi.org/10.1159/000065672

Appendix 1: I can hear my brain booklet

I Can hear my brain

Cognitive Neuroscience Lab

University of
South Australia

This is Willis.

Willis likes to play games and watch television.

Today is Willis' first visit to the UniSA Lab.

First Willis puts on the magic hat.

Then we put some special gel on both sides of his head to make the special ears work.

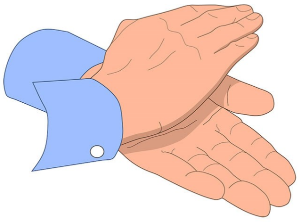
Next, the special ears are attached to the magic hat!

Now Willis is ready to watch videos!

We use our brains to do many things. They help us...

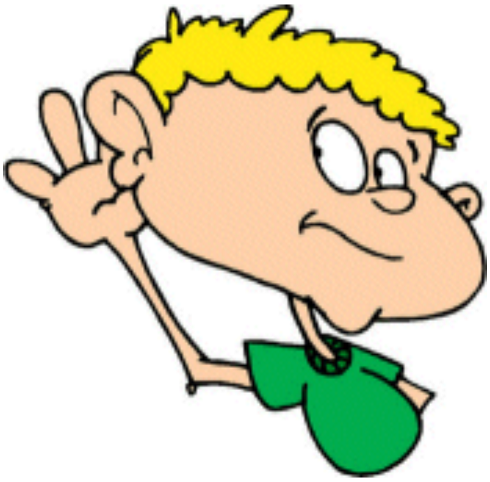
Play,

Dance



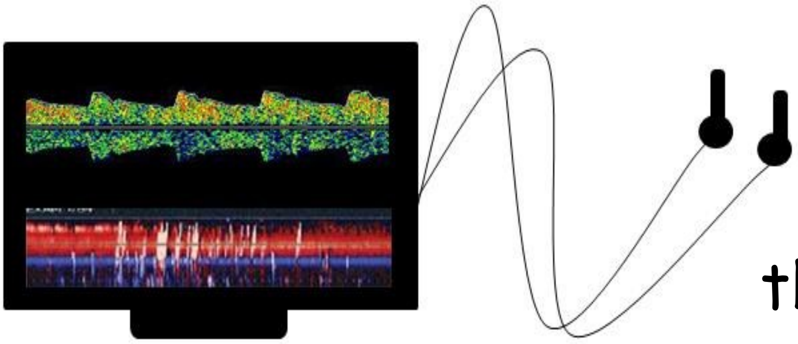
And Clap.

To help us do these things our brain needs special food. Our blood carries this food to our brain.



This makes a very
special quiet
sound!

To hear this sound, we wear a magic hat!
Then, we put on some special ears.



The sound travels through the special ears, along a wire and to our computer!

There are two special ears, for each side of the brain.

“What’s this?”

“It’s a Baby”

So, when someone says the name of a picture, we can hear which side of the brain makes the loudest noise!