Citation: Jones, P. R. ORCID: 0000-0001-7672-8397 (2018). The development of perceptual averaging: Efficiency metrics in children and adults using a multiple-observation sound-localization task. Journal of the Acoustical Society of America, 144(1), pp. 228-241. doi: 10.1121/1.5043394

TITLE

The development of perceptual averaging: efficiency metrics in children and adults using a multiple-observation sound-localization task

RUNNING TITLE

The development of perceptual averaging

AUTHORS

Pete R. Jones

Institute of Ophthalmology, University College London (UCL), 11-43 Bath Street, London EC1V 9EL; p.r.jones@ucl.ac.uk

ABSTRACT [Max 200 Words; Currently 194]

This study examined the ability of older children to integrate spatial information across sequential observations of bandpass noise. In Experiment I, twelve adults and twelve 8—14-year-olds localized 1—5 sounds, all presented at the same location along a 34° speaker array. Rate of gain in response precision (as a function of N observations) was used to measure integration efficiency. Children were no worse at localizing a single sound than adults, and --- unexpectedly --- were no less efficient at integrating information across observations. Experiment II repeated the task using a Reverse Correlation paradigm. The number of observations was fixed ($N = 5$), and the location of each sound was independently randomly jittered. Relative weights were computed for each observation interval. Distance from the ideal weight-vector was used to index integration efficiency. The data showed that children were significantly less efficient integrators than adults: only reaching adult-like performance by around 11 years. The developmental effect was small, however, relative to the amount of individual variability, with some younger children exhibiting greater efficiency than some adults. This work indicates that sensory integration continues to mature into late childhood, but that this development is relatively gradual.

KEY WORDS: *integration efficiency, multiple observations, sound localization, reverse correlation*

I. INTRODUCTION

On simple psychophysical tasks, older children often perform as well as adults[1]. For example, the ability to discriminate the frequency of two tones is adult-like by around 8 years of age[2], while the ability to localize a single sound matures by around 6 years[3]. In everyday life, however, we are often presented with complex scenes, containing multiple sources of stochastic information. In such circumstances, perceptual judgments are limited not only by our ability to encode individual stimuli, but also by our ability to integrate multiple observations together, to make a single, overall decision.

Outside of audition, children's ability to integrate information across multiple sensory 'channels' is believed to remain immature until late childhood. For example, children up until 10 – 12 years have been shown to fixate disproportionately on a single modality in multisensory tests of navigation[4], visuohaptic size discrimination[5], and audiovisual stimulus detection[6] (for reviews, see [[7],[8]]). While within vision, the ability to combine different stimulus features (e.g., texture and stereoscopic disparity) to judge depth has been found to mature only by around 11-12 years[9],[10]. Within audition, the developmental time course is unknown. However, there is clear evidence of suboptimal integration in early childhood. For example, Allen, Jones, & Slaney (1998)[11] observed that adults exhibited a substantial benefit (~8 dB) on a tone-in-noise detection task when the target was positioned spectrally off-center. In contrast, preschool children (4--5 years) gained no such benefit, indicating that they were unable to exploit both pitch and level cues.

It is also striking that where the development of sensory integration has been studied, it is often limited to tasks involving only two channels of information. And it is known that as the number of channels increases, even adults' performance start to deviates from the ideal[12–14] -- possibly due to constraints on memory or attention. This raises the possibility that, in arguably more realistic scenarios, where more than two sources of information are present, children may not be any poorer than adults at integrating information. Indeed, one recent study by Leibold and Bonino[15] suggests this might be the case. There, it was found that children's detection thresholds for a tone in noise improved progressively the more the target was repeated (N = 1 to 5), and the rate of improvement did not differ significantly between children and adults.

The purpose of the present study was to quantify the ability of older children (aged 8 – 14 years) to integrate sequential auditory signals, and to determine at what age this ability matures. To quantify efficiency, we used a 'multiple observation'[12] perceptual averaging task. On each trial, the listener was presented with a sequence of sounds, all centered on a single location along the azimuth (location randomized between trials). The listener's task was to listen to all N sounds, before judging the (single) source location. Two separate techniques were used, in two independent experiments, to estimate the efficiency with which listeners combined the N observations to form a single estimate of location. Each experiment is reported more fully in turn, but in brief:

Experiment I measured integration efficiency using a relatively old method based on the rate of gain in response precision as a function of N observations. During the experiment, N was varied randomly between 1 to 5. Within a single trial, all N sounds were presented at the exact same location. This meant that every observation was equally informative, and the response precision of the ideal observer are predicted to improve at a rate of $\sqrt{N}$[16]. To the extent that listeners failed to integrate additional observations, their response precision would improve at a lesser rate. The rate of gain provided an index of integration efficiency.

Experiment II used a more modern measure integration efficiency based on Reverse Correlation. The number of observations was fixed at N = 5 and the location of each sound was randomly jittered between observations. Each of the five observations therefore predicted a slightly different response. The relative correlation between the listener's actual responses, and the predicted responses for each of the five temporal intervals, therefore provided a measure of the *relative weight* given to each observation. To the extent that the listener utilized all five observations, equal weight should be given to each. Conversely, a suboptimal integrator would over-weight some temporal intervals, and under-weight others. The similarity of the observed weights vector to the ideal provided an index of integration efficiency.

Previous studies have used variants of both methods in adults[12],[13]. These studies have shown that adults are effective but sub-optimal integrators: deriving a measurable benefit from every additional information channel, but less benefit than would be predicted by an ideal observer. The novel aspect of this present work was the application of these methods to children. It was therefore unknown how they would perform. In particular, it was unknown: how children's efficiency compared to adults, and which (if any) of the N observations children would fail to exploit.

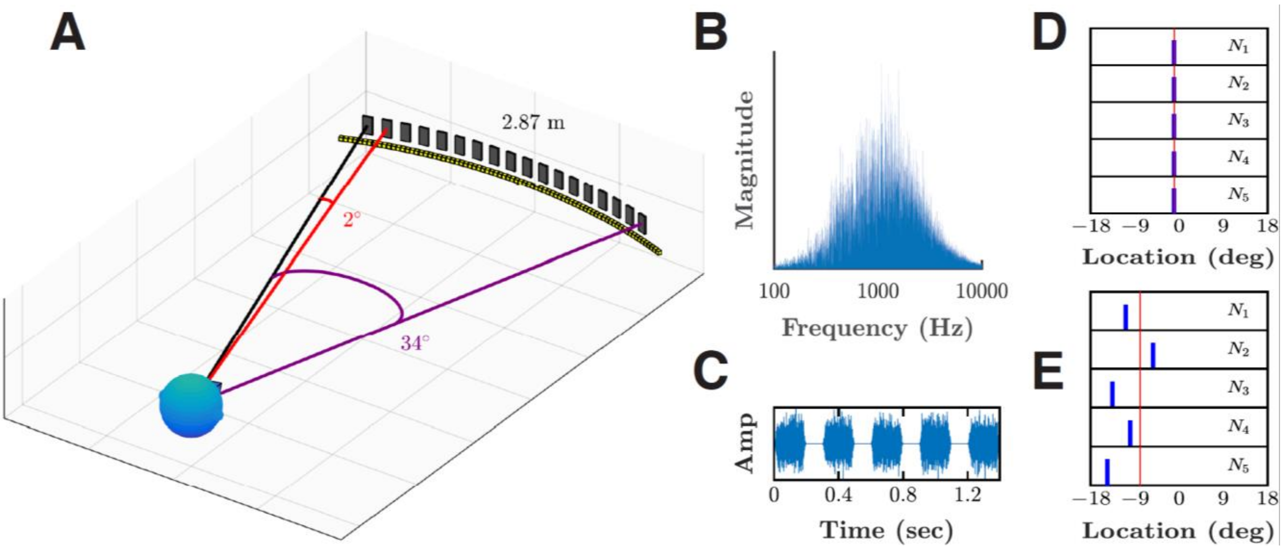


FIG 1. Stimuli and test apparatus for both experiments. ***(A)*** The listener's task was to locate the [single] source location of N noise bursts. Stimuli were presented along the azimuth, using 18 speakers distributed uniformly at 2° intervals along a 34° arc. Eighty LEDs arranged below the speakers were used for response-input, feedback, and fixation-cuing; ***(B)*** Each observation consisted of a 200 ms band-passed noise burst (1 octave bandwidth), centered at 1 kHz. ***(C)*** Each trial consisted of N observations (shown here: $N = 5$), presented sequentially with an inter-stimulus interval of 100 ms. ***(D)*** In Experiment I, N varied from 1 to 5, between blocks, in random order. Within each trial, the target location (thin red vertical line) varied randomly, and all sounds (thick blue lines) were presented at the target location (shown here: target = -1.25°). ***(E)*** In Experiment II, N was fixed at 5, and the location of each sound was randomly distributed around the target location, based on independent samples from a truncated-gaussian random variable (shown here: target = -9.25°).

II. EXPERIMENT I: Relative gain in response precision as a function of *N* observations

The goal of Experiment I was to quantify integration efficiency in children and adults, using the relative gain in response precision as the number of observations, *N*, increased. The logic of this method is derived from basic Signal Detection Theory[12], and is described more fully elsewhere[12]. In brief: let us assume that the response to a single sound is determined by some putative 'internal response', which is a scalar value proportional to the observed stimulus value, plus a sample of additive noise (i.e., due to random error due to intrinsic neuronal, physiological, or cognitive variability): $x + \varepsilon$. And let us model the additive noise term as a zero-mean Gaussian variable, $\epsilon \sim \mathcal{N}(0, \sigma_{int}^2)$ – a choice that is mathematically expedient, but which in the present case is also supported by the empirical data (see Fig S1 in the Supplementary Material[17]). If we operationalize response precision as the reciprocal of the standard deviation of the observed response error, $\frac{1}{\sigma}$, then response precision in the single stimulus condition is determined purely by the standard deviation ('magnitude') of the internal noise, σ_{int}:

$$\text{PRECISION}_1 = \frac{1}{\sigma_1} = \frac{1}{\sigma_{int}} \qquad \textbf{(Eq 1)}$$

When presented with multiple, equally-reliable observations, the ideal observer will mean-average the *N* internal responses: $\sum_{i=1}^{N}[x_i + \varepsilon_i]$. The decision variable will therefore be the mean of *N* normally distributed random variables, which is itself a normally distributed random variable with a mean of $\bar{x}$ and a standard deviation of $\sigma/\sqrt{N}$. We would therefore expect the response precision of an ideal observer to improve at a rate of $\sqrt{N}$ (for more detailed theory, see References [[12],[16]]).

Conversely, a listener who used only some proportion, *k*, of the additional information, would gain proportionally less benefit from observing additional observation, thus:

$$\text{PRECISION}_N = \frac{1}{\sigma_N} = \frac{1}{\sigma_{int}/\sqrt{1+k(N-1)}} = \frac{\sqrt{1+k(N-1)}}{\sigma_{int}} \quad \textbf{(Eq 2)}$$

For example, when $k = 0$, precision with N observations would be the same as precision with one observation (no improvement). As k increases towards 1, the rate of relative improvement becomes closer to the ideal: $\sqrt{N}$. Thus, if $N = 3$ and $k = 0.5$, precision would be ~1.41 ($\sqrt{2}$) times greater than precision given a single observation, while if $k = 1$ precision would improve by ~1.73 ($\sqrt{3}$).

By combining Eqs 1 and 2 it can be seen that σ_N/σ_1 (the ratio of response precision given N observations, to precision given one observation only) is determined solely by the single unknown parameter k, together with the experimentally controlled parameter N:

$$\frac{\text{PRECISION}_1}{\text{PRECISION}_N} = \frac{\sigma_N}{\sigma_1} = \frac{\sigma_{int}/\sqrt{1+k(N-1)}}{\sigma_{int}} = \frac{1}{\sqrt{1+k(N-1)}} \quad \textbf{(Eq 3)}$$

Thus, by plotting empirical values of σ_N/σ_1 as a function of N, the best-fitting value of k (proportion of observations used) can be estimated. This is illustrated graphically in Figure 2, which shows individual data for two individuals, superimposed against isobars for various values of k, ranging from no integration ($k = 0$) to full integration ($k = 1$). By inspection, it can be seen that one listener (red circles) used only ~50% of the additional information, while a second listener (blue diamonds) was a near-optimal integrator (~100%). In practice, values of k were estimated numerically by finding the value of k that minimized the least-square error between Eq 3 and the empirical data.

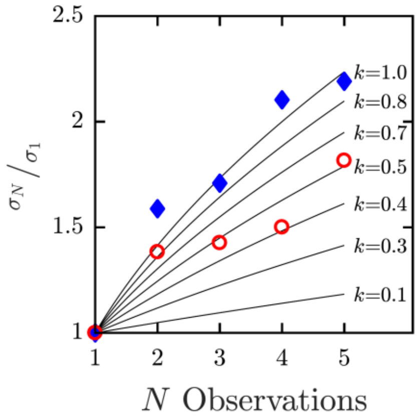


FIG 2. Experiment I: The determination of k (proportion of observations used), using five successive observations of a 1-octave noise burst. Black lines are isobars denoting the rate of gain predicted as integration varies from k = 0 (no integration) to k = 1 (full integration. Red circles and blue diamonds are data from two individual listeners.

A. Experimental Methods

1. Task Overview

As illustrated in Figure 1, the task was to localize the [single] source of *N* noise bursts (‘observations’), where *N* varied from 1 to 5 between blocks (random order). The *N* observations were presented sequentially at a random location along a 34° array of loudspeakers, which was arranged in a frontal arc around the participant. After all *N* observations, the participant made a single response, by using a rotary dial to position a light at the perceived sound-source location. Participants were encouraged to “listen carefully to all of the sounds without moving your head, before deciding where the sounds were coming from”.

2. Participants

Participants were 12 normal hearing children, aged 7.9 – 13.9 years (μ = 11.0, σ = 2.0), and 12 normal hearing adult controls, aged 18 – 30 years. Adults were recruited through the UCL Psychology Subject Pool (‘SONA’), and received £7.5/h compensation. Children were recruited through the UCL Child Vision Lab volunteer database, and received certificates and small toys. Written consent was obtained from all participants (adults) or the responsible caregiver (children). Children themselves also gave written assent. The experiment was conducted in accordance with UCL Research Ethics Committee approval (#7611/001).

3. Stimuli & Apparatus

Each stimulus consisted of *N* band-pass noise bursts separated by inter-stimulus intervals of 100 ms. Each noise burst was 200 ms in duration, including 10 ms $\cos^2$ on/off ramps (see Fig 1B-C). Each burst was independently randomly generated by filtering white Gaussian noise through a pair of second-order Butterworth band-pass filters, with cut-offs 1-octave either side of 1 kHz (i.e., 0.5 kHz High Pass, 2 kHz Low Pass). Stimuli were presented over loudspeakers, at an intensity of 59.5 to 60.5 dB SPL. The small amount of level jitter was drawn randomly from a uniform distribution, and was designed to prevent loudness inadvertently becoming a location cue (e.g., due to errors in calibration, or systematic differences in room-acoustics).

The exact choice of stimulus is not expected to have influenced the ability of children or adults to integrate observations. However, the bandwidth of the signal (1 octave) was significant from a practical perspective. The ability of listeners to localize sounds stimuli declines precipitously for narrower bandwidths[18], and it was observed during piloting that listeners often became unmotivated when presented with narrowband noise or pure tones. In such circumstances, listeners were also liable to be influenced in their responses by *a priori* information (i.e., the visible extent of the speaker ring). Very wideband stimuli were also deemed inappropriate, as, consistent with previous findings[18], some pilot listeners performed close to ceiling when presented with a single burst of white noise at certain locations. The center frequency of the stimulus (1 kHz) meant that the signal contained both ITD and ILD cues. However, the choice of center frequency is unlikely to have affected observed behavior substantially, as the ability to localize broadband stimuli along the azimuth is largely independent of center frequency for bandwidths of 1 octave or greater[18].

Stimuli were presented using an array of eighteen speakers (Visaton SC 5.9; Visaton GmbH, Haan, Germany), which were positioned symmetrically, equidistant from the listener. The speakers were uniformly-spaced in 2° intervals along a circular arc spanning ±17° either side of the listener's midline [Fig 1A]. Each speaker was located 2.87m from the listener. To allow sounds to be located continuously anywhere along the 34° arc, Vector Distance Panning was used to interpolate between speakers[19]. Panning was used to ensure that the distribution of target locations was as close to gaussian-distributed as possible, and also to minimize the possibility that listeners might learn the *N* discrete speaker locations. The use panning may have introduced a small amount of additional variability into listeners' location judgments. However, performance was similar to previous studies in which panning was not employed (see General Discussion). An acoustically transparent curtain was arranged in front of the speakers, to prevent listeners from assuming that sounds were only ever located at the 18 discrete speaker locations.

Stimuli were digitally synthesized in MATLAB v7.4 (2012a, The MathWorks, Natick, MA) using a sampling rate of 44.1~kHz and 24-bit quantization. Stimulus presentation was controlled using the Psychophysics Toolbox v3[20] ASIO wrapper (Steinberg Media Technologies, Hamburg). Digital-to-analogue conversion was carried out by a Focusrite Saffire PRO 40 (Focusrite plc, UK) external sound card (channels 1 to 10), and by an Ultragain Digital ADA8000 (Behringer GmbH, Willich, Germany) ADAT interface (channels 11 to 18). Audio signals were amplified using nine Lvpin Hi-Fi 2.1 stereo amps (Lvpin Technology Co. Ltd, Suzhou, China). Output levels were equalized using an Investigator 2260 sound level meter (Brüel & Kjær, Nærum, Denmark), and were adjusted to ensure no noticeable differences in intensity or timbre.

Directly below the speakers was an array of 80 light-emitting diodes (12 mm diffused digital LED pixels; Adafruit Industries, New York, New York, USA), distributed uniformly between ± 19.75°, in intervals of 0.5°. The LEDs were used to provide: (i) a central fixation-target prior to each trial, (ii) post-trial feedback on the true target locations, and (iii) the means by which observers responded (see Procedure, below). An Arduino Uno microcontroller (SmartProjects, Strambino, Italy) was used to interface between the control computer and the LED pixels (see Reference [21]). When making responses, the listener controlled which one of the 80 LEDs was illuminated by rotating a dial (PowerMate USB; Griffin Technology, Nashville, Tennessee, USA). The participant used a keystroke to indicate when done, at which point their response was logged.

With both children and adults, the experimenter was present throughout testing, to provide instruction and encouragement. A minority of the children were accompanied by a caregiver (generally their parent), who sat outside the child's field of vision and who was asked to remain silent during testing.

4. Procedure

Each trial commenced with a 660 ms visual fixation target, during which the two central LEDs (±0.25°) were illuminated bright red. *N* successive 200 ms noise bursts were then presented at the target location, separated by inter-stimulus intervals of 100 ms. The target location was randomly selected on each trial, using a uniform distribution between ± 16.75°, rounded to the nearest 0.5° to ensure that the target always fell directly above one of the LEDs (i.e., to ensure accurate responses and veridical feedback). In instances where the target fell between two speaker locations, panning was used to present the stimulus, as described above (Stimuli & Apparatus).

Following stimulus presentation, the listener responded by 'pointing' to the perceived sound source location. To do this, one of the two central LEDs was randomly selected and was illuminated white. The listener was then given unlimited time to 'move' this light to the perceived sound-source location, using a rotary dial to control which of the LEDs was illuminated. Feedback was then given in the form of a green LED light, which was presented at the target location for 660 ms.

The test session consisted of 250 trials, divided equally between five conditions: *N* = {1, 2, 3, 4, 5}. Each condition was tested in a separate block of 50 trials, and the order of the blocks/conditions was randomized between listeners. After each block, the listener was given the opportunity to take a short break, as required. Each listener completed a single session, which lasted approximately 60 minutes (including consenting, practice, and breaks).

Before the test trials, each listener completed five practice trials. These trials were identical to the test trials, and were all drawn from the *N* = 3 condition. During this period, the listener was encouraged to listen carefully to all the sounds, before deciding where [all] the sounds were coming from.

B. Results

Figure 3 shows mean response precision for adults and children. To analyze these data, a 5x2 mixed ANOVA was performed with a within-subject variable of *N OBSERVATIONS* (5 levels: N = 1--5), and a between-subject variable of *AGE* (2 levels: children, adults). There was no significant main effect of *AGE* [$F_{(2,22)} = 1.37$, $p = 0.255$, *n.s.*], indicating that children were no less precise than adults in terms of their overall localization ability (although, *prima facie*, a possible trend towards higher precision in adults is apparent in Fig 4). In particular, an independent-samples *t*-test indicated that children were not significantly less precise than adults in the $N = 1$ condition [$t_{22} = 1.38$, $p = 0.183$, *n.s.*].

However, there was a clear main effect of *N OBSERVATIONS* [$F_{(4,88)} = 7.14$, $p < 0.001$], indicating that precision improved as the number of observations increased. This implies that at least *some* integration was taking place. Accordingly, precision in the $N = 5$ condition was significantly higher than in the $N = 1$ condition, both for children [Paired t-test: $t_{11} = 3.80$, $p = 0.003$], and adults [$t_{11} = 3.79$, $p = 0.003$]. There was no interaction between *AGE* and *N OBSERVATIONS* [$F_{(4,88)} = 0.20$, $p = 0.937$, *n.s.*], suggesting that the rate of improvement, and therefore the amount of integration, was similar between age groups.

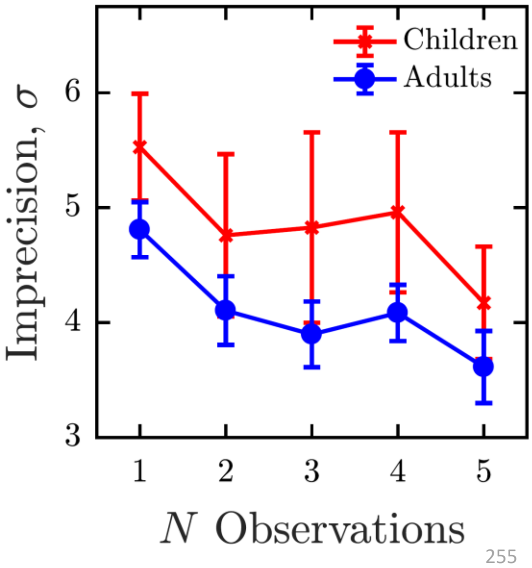


FIG 3. Experiment I: Group-mean [± 1 S.E.] response variability for children (red crosses) and adults (blue circles), shown as a function of *N Observations*. Lower values denote greater precision. For the ideal observer, imprecision would be expected to decrease at a rate of $\sqrt{N}$.

The foregoing implies that both children and adults integrated information from at least two observations (in the nomenclature of Boyaci and colleagues[22], adults and children were both 'effective integrators'). However, these analyses do not allow us to quantify the relative efficiency of children and adults.

To formally assess integration efficiency, we computed σ_N/σ_1 and estimated k (proportion of observations used), using the procedure described in the Methods. Results are shown for individuals in Figure 4. By inspection, there was substantial inter-individual variability, but no systematic difference between children and adults. This was confirmed statistically using a Mann-Whitney U test, which found no significant difference in efficiency, k, between children and adults [$U = 148$, $Z = -0.09$, $p = 0.931$]. In short, neither age group appeared better at integrating sensory information [Fig 5].

A Wilcoxon Signed-Rank test indicated that, on average both children [$p < 0.001$] and adults [$p < 0.001$] deviated significantly from the ideal observer (dashed lines in Figs 4 & 5), indicating that both were suboptimal, and failed full use of the additional information. However, it can be seen in Figure 4 that there were individual exceptions, with some adults and some children performing close to the ideal.

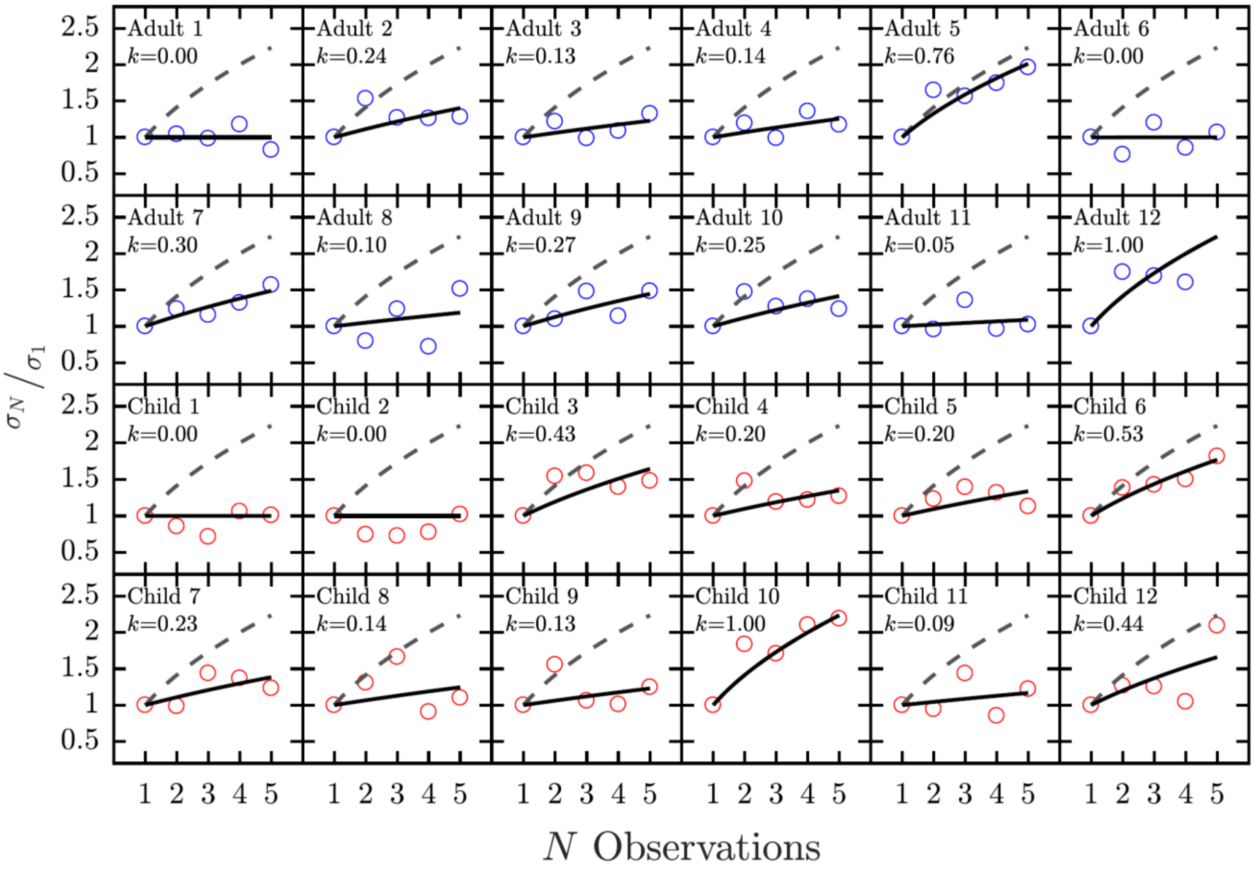


FIG 4. Experiment I: Value of σ_N/σ_1 for all individuals. Solid lines represent least-square fits of Eq 3 to the data, from which estimates of the integration index, k, were derived (see Fig2 for details). Dashed lines show the ideal rate of gain ($\sqrt{N}$). Individual children have been ordered by age (ascending).

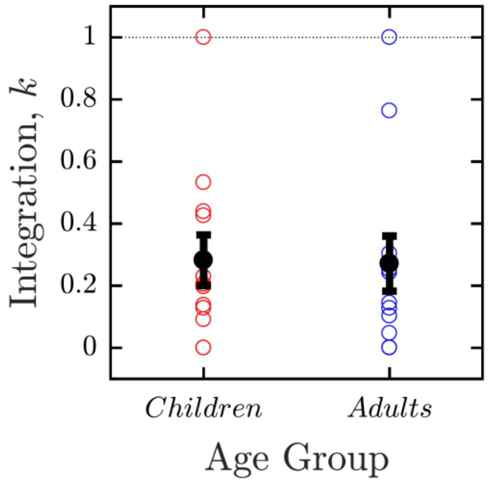


FIG 5. Experiment I: Group-mean [± 1 S.E.] integration efficiency for children and adults (same data as Fig 4). Markers indicate values of *k* for individual subjects. Horizontal dashed line represents the ideal observer.

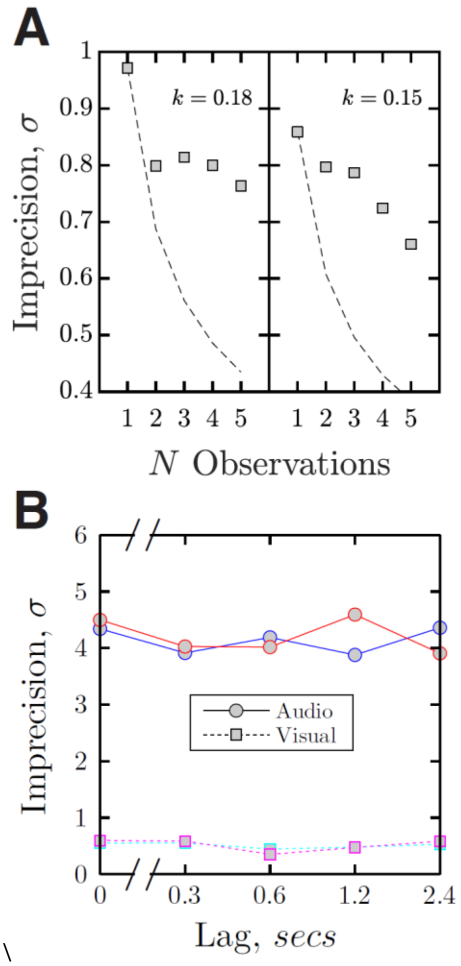
C. Interim Discussion

The results from Experiment I showed that both children and adults are able to integrate information across multiple, sequential observations. However: (i) both children and adults were suboptimal, and on average exhibited lower integration efficiency than the ideal observer (although substantial individual variability was observed). Furthermore, and contrary to expectations: (ii) children were, on average, no less efficient at integrating information than adults.

The fact that integration efficiency was relatively low in adults stands in apparent contradiction to the wider 'cue-combination' literature, where sensory integration in adults is generally reported to be near-optimal (for a review, see [23]). However, findings of near-optimality are generally predicated on tasks involving only two channels of information. In contrast, when, as in the present task, larger numbers of channels are presented sequentially, studies in both vision[13],[14] and audition[12] have, like the present work, tended to report effective but suboptimal integration.

That children's localization precision improved at the same rate as adults is consistent with a study by Leibold and Bonino (2009)[15], where children's detection thresholds for a repeated-tone in noise were found to improve at the same rate as adults (see Introduction). Furthermore, the pattern of results observed in Figure 4 are also reminiscent of data from He, Buss, & Hall (2010)[24], in which children were asked to detect brief pure tones embedded in a continuous bandpass noise. As the duration of the target tone increased, detection thresholds improved. And although thresholds were consistently poorer for children than adults, the rate of improvement was similar for younger children (5 – 7.5 years), older children (7.5 – 10 years) and adults. The absence of any developmental effects in the present experiment were, nonetheless, unexpected, given the overwhelming consensus in the wider developmental literature that sensory integration remains immature until ~11 years[7–10].

The conclusions of Experiment I are, however, open to question. To see why, note that by inferring efficiency from the relative gain in response precision, we are assuming, implicitly, that all internal noise is occurs ‘early’ in the encoding process, in the sense that it arises independently in the peripheral auditory system, before any sensory observations are integrated, and so will cancel-out across repeated observations[25]. In contrast, there are many potential sources of response imprecision that are irreducible, and liable not to cancel-out across observations. For example, motor noise, memory decay, key press errors, variations in response criterion, sensory noise that is correlated across observations, interference between sensory observations (e.g., masking), and/or difficulties in mapping between auditory (stimulus) space and visual (response) space, may all add noise to the listener’s responses, and do so in a way that does not decrease with N (or may even increase). Of these, some potential sources of irreducible noise can be discounted by simple control experiments. For instance, when the experiment was repeated using a visual location cue, overall imprecision was greatly reduced, but continued to decline as a function of N (Fig 6A). This demonstrates that irreducible motor noise is unlikely to be primary limiting factors in the main experiment. Similarly, in a small number of adult controls, imprecision was found not to vary significantly when the lag between a single stimulus and response was systematically increased, either when using a visual (Fig 6B squares) or auditory (Fig 6B circles) stimulus. This suggests that simple memory-decay is also unlikely to be a limiting factor in the main experiment. Other forms of irreducible noise cannot, however, be ruled out.



\

FIG 6. Experiment I control data, from six additional adults. These controls did not participant in the main experiment and were naïve to the task (***A***) Data from a visual localization task. The task was identical to the main experiment, except that the *N* noise burst were replaced with *N* pulses of white light. As in the main experiment, indices of integration efficiency, *k*, were computed using Eq 3. The values of *k* are comparable with those for the main auditory task (Figures 4 & 5). (***B***) Control data for an *N*=1 localization condition in which a temporal lag was interposed between stimulus presentation and the participant's response. Participants were instructed to keep fixating centrally until the response light appeared. Stimuli consisted of either sounds (circles) or lights (squares). Each colored line represents a different observer.

To see why irreducible is problematic, note that without the common/convenient assumption that all internal noise is reducible, Equation 2 becomes:

$$\mathrm{PRECISION}_N = \frac{1}{\sigma_N} = \frac{1}{\sqrt{\sigma^2_{int-r}/[1+k(N-1)] + \sigma^2_{int-ir}}} = \sqrt{\frac{1+k(N-1)}{\sigma^2_{int-r} + \sigma^2_{int-ir}[1+k(N-1)]}} \quad \textbf{(Eq 4)}$$

where σ_{int-r} and σ_{int-ir} are the reducible and irreducible internal noise components, respectively. It follows that Equation 3 becomes:

$$\frac{\sigma_N}{\sigma_1} = \frac{\sqrt{\sigma^2_{int-r} + \sigma^2_{int-ir}}}{\sqrt{\sigma^2_{int-r}/[1+k(N-1)] + \sigma^2_{int-ir}}} \quad \textbf{(Eq 5)}$$

The key point to note is that, unlike Equation 3 (which was used to fit the data in Figures 4 and 5), the internal noise terms in Equation 5 no longer cancel out. The ratio σ_N/σ_1 therefore no longer provides an unambiguous measure of integration efficiency, k. Thus, with the model expressed by Equation 5, Listener A may show a greater rate of improvement than Listener B *either* because Listener A is a more efficient integrator ($k_A > k_B$), or because a greater proportion of Listener B's internal noise is irreducible $\left(\left[\frac{\sigma_{int-ir}}{\sigma_{int-r}}\right]_A < \left[\frac{\sigma_{int-ir}}{\sigma_{int-r}}\right]_B\right)$.

The two key corollaries of this is that we cannot be sure that children are as efficient as adults (i.e., since the proportion of irreducible noise may change with age), and we cannot be sure that individual listeners --- either children or adult --- were in fact integrating suboptimally. To the extent that internal noise is irreducible, listeners may be better integrators than the results of Experiment 1 suggest, and the estimates of k reported in Figure 4 and 5 are only lower bounds on integration efficiency.

One way to address the problem of irreducible noise is to explicitly introduce additional external noise that we know to be reducible. For example, Swets et al (1959)[12] performed a multiple-observation tone detection task analogous to the localization task reported here. They similarly found that adult performance improved as a function of N, and that the rate of gain was relatively small. Notably though, they also ran a second condition in which independent samples of external noise were added to each observation. In that case, the rate of gain improved markedly, and was close to optimal ($\sqrt{N}$) for most listeners. This suggests that if Experiment I were repeated with external noise added, estimates integration efficiency might increase, and may start to differ between children and adults. Furthermore, since any external noise is directly observable, it also becomes possible to perform trial-by-trial ('molecular'[26]) analyses, to determine which observations the listener predicated their response upon (see Experiment II). In this way, it is possible to characterize not just whether, but in what way integration is suboptimal. This is the approach taken in Experiment II.

III. EXPERIMENT II: Relative decision weights using Reverse Correlation

The goal of Experiment II was to again quantify integration efficiency in children and adults. This time, however, external noise was added to each observation, and a Reverse Correlation technique was used to estimate each listener's decision strategy.

The Reverse Correlation methodology is described in detail elsewhere[26]–[28], and has been used previously in adults to study integration of sequentially presented visual stimuli[13],[14]. In brief: just as in Experiment I, *N* noise bursts were presented on each trial, and the listener was asked to make a single judgment of location. However, the location of each individual noise burst was independently randomly jittered prior to presentation, such that each observation predicted a slightly different response (Fig 1E). By comparing the listener's trial-by-trial responses (irrespective of their accuracy) to the predictions of the various observations, one can estimate the relative degree to which the listener attends-to/relies-upon each observation. In practice, this procedure was carried out in the present study using a multiple regression model[27] (MATLAB's GLMFIT routine).

The result of this analysis is a vector of estimated relative weights, ω_{est}, where the i^{th} weight indicates the listener's relative reliance on the i^{th} observation. By convention we shall normalize this vector such that the absolute magnitudes sum to 1. For example, a listener who only used the first observation would exhibit relative weights of ω_{est} = [1 0 0 0 0]. Conversely, when, as in the present case, all 5 observations are equally informative, the ideal weight vector, ω_{idl}, is: [0.2 0.2 0.2 0.2 0.2].

The deviation of the observed weights, ω_{est}, to the ideal, ω_{idl}, provides an index of integration efficiency, η_{ω}, which we can formalise in terms of root-mean-square error[29]:

$$\eta_{\omega} = 1 - RMS = 1 - \sqrt{\frac{1}{N}\left(\sum_{i=1}^{N}\left[\omega_{est}(i) - \omega_{idl}(i)\right]^2\right)} \qquad \textbf{(Eq 6)}$$

Thus, $\eta_{\omega} = 1$ represents perfect efficiency, and lower values indicate a progressive loss of sensory information. Note that this integration index is not directly comparable to the value k, reported previously in Experiment I, although conceptually both are intended to capture the degree to which listeners are able to exploit multiple observations.

Crucially, the external noise was sampled independently for each observation, and so would cancel out across observations. This guaranteed that listeners would be more precise when integrating across observations, thereby swamping the effects of any irreducible internal noise. Furthermore, with this method of analysis, some forms of irreducible noise, such as motor error, are largely partialled out from the estimate of integration efficiency, since they add noise to the final response, but in a way that would not be expected to affect the estimated weight-vector, ω_{est} (i.e., motor noise would not systematically bias responses towards any single observation interval).

A. Experimental Methods

1. Task, Stimuli, Apparatus & Procedure

The task was identical to Experiment I, with two exceptions. Firstly, the number of observations was fixed at $N = 5$ for every trial (to ensure sufficient data for the Reverse Correlation analysis). Secondly, to facilitate the Reverse Correlation analysis, external noise, in the form of truncated Gaussian jitter, was added independently to every stimulus, prior to presentation. This jitter needed to be large enough that, across trials, each observation predicted a measurably different vector of responses, but small enough that listeners did not come to suspect that some observations were unreliable. To this end, the jitter was determined by a zero-mean truncated Gaussian distribution, with a standard deviation of 3°, and a min/max of ±7° (i.e., 2.333σ). These parameters ensured that stimuli would not fall far outside the range of error predicted by internal noise alone (see Fig S1 in the Supplementary Material), and when questioned after testing, participants did not report being aware of the external noise manipulation. To further prevent stimuli falling outside the total span of speakers, the target location (i.e., the center of the Gaussian distribution) was limited to the central ±10° of the speaker arc. Jittered locations were not rounded to the nearest LED location and, unlike Experiment 1, the weighted-average location of the five observations was not guaranteed to fall directly above a target LED. This may have introduced a small amount of quantization error into listener's responses, but this not expected to have had any effect on the reported findings. Each participant completed four blocks of 50 trials (all $N = 5$), in a single session lasting approximately 60 minutes (including breaks).

2. Participants

A new cohort of participants was recruited, consisting of 12 normal hearing children, aged 8.3 – 13.9 years ($\mu = 10.1$, $\sigma = 1.7$), and 12 normal hearing adult controls, aged 18 – 30 years. None of the listeners from Experiment I participated, and there was no significant difference in the age of the children versus their Experiment I counterparts [$t_{22} = 1.22$, $p = 0.24$, *n.s.*].

B. Results

We begin by considering the data for each individual listener, shown in Figure 7. To the extent that an overall pattern can be discerned, the general trend was towards response strategies that prioritized the first (primacy) or last (recency) observation. However, there was considerable individual variability in both response strategy and overall efficiency. Thus, while Adult 13 and Child 14 both up-weighted the first/last observation, and down-weighted the central observation, Adult 17 exhibited the inverse pattern: relying predominantly on the 3rd observation, and relatively little on the first/last observations. Only one listener (Child 20) appeared to base their responses on only a single observation. However, few listeners approximated the ideal -- though even in this respect were exceptions (cf. Adult 19, Adult 24, Child 15). Individual variability in weight efficiency, η_{ω}, was positively correlated with response precision [Pearson's linear correlation: $r_{22} = 0.58$, $p = 0.003$] – with more efficient weightings associated with lower response variability. This suggests that the reverse correlation method reliably captures performance-relevant integration strategies.

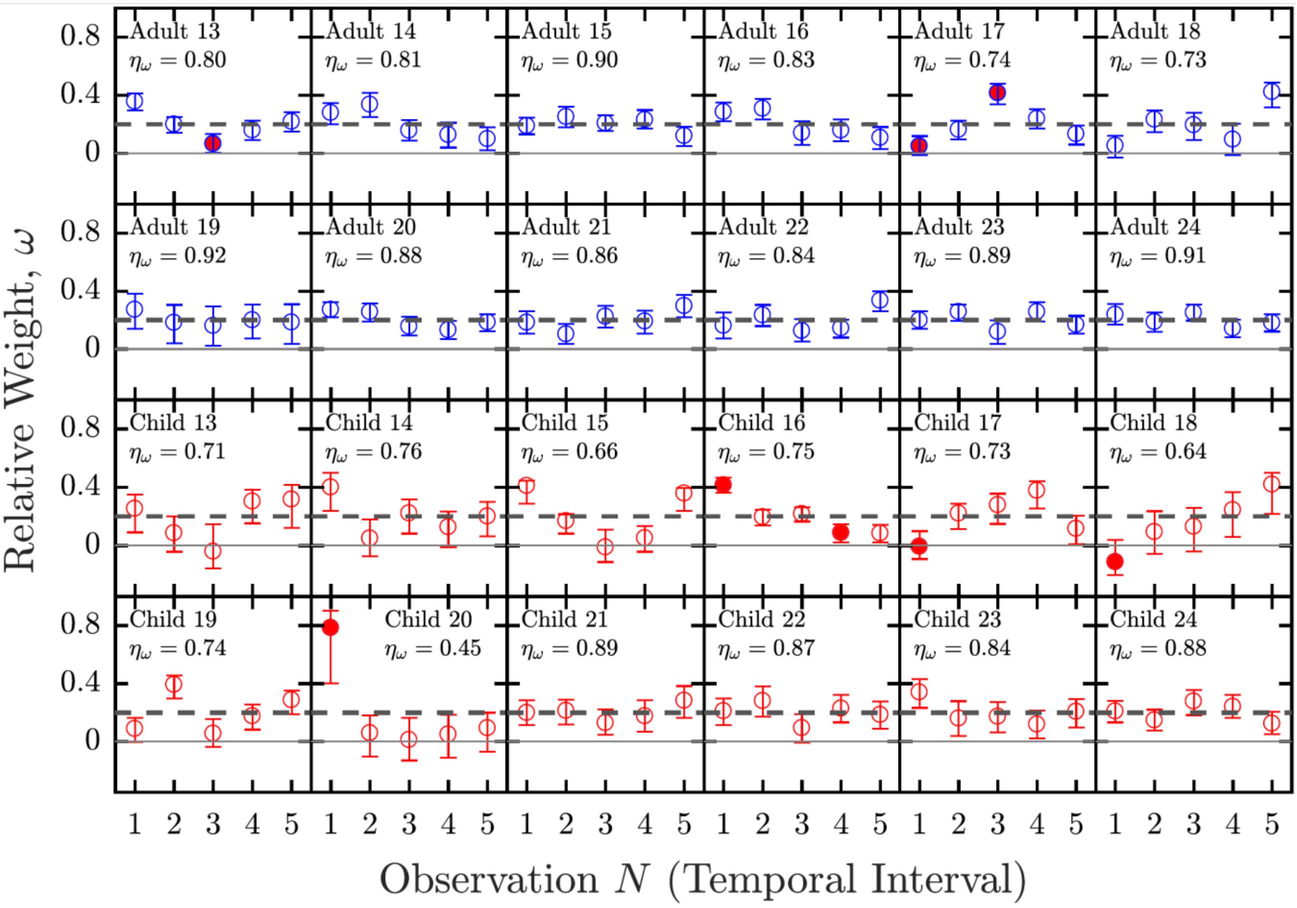


FIG 7. Experiment II: Relative weight vectors for all individuals, with bootstrapped 95% standard error bars. Dashed lines show the ideal weight vector. Shaded markers denotate instances where empirical weights deviated significantly from the ideal. Individual children have been ordered by age (ascending).

A significant difference in integration efficiency, η_ω, was observed between children and adults [$t_{22} = 2.49$, $p = 0.021$], with adults tending to exhibit more efficient decision strategies [Fig 8A]. To confirm that this difference was not due to one poor performing child (see Fig 8A), this analysis was also repeated with this individual excluded [$t_{21} = 2.33$, $p = 0.030$], and using a non-parametric analog [*Wilcoxon rank sum*; $Z = 2.17$, $r = 0.44$, $p = 0.030$]. In both cases, the same age-difference was found. Both children [$t_{11} = -6.50$, $p < 0.001$] and adults [$t_{11} = -8.29$, $p < 0.001$] differed significantly from the ideal observer [horizontal dashed line] – indicating that, on average, both age-groups were suboptimal.

To examine the developmental time-course, Figure 8B plots integration efficiency as a function of age. Based on the best fitting broken-stick function, it appears that adult-like performance was reached by 11.4 years. However, even many younger children fell within the 95% population limits of the adults (Fig 8B, shaded region). Furthermore, the fitted curve only explained 44% of the variability in the raw data ($R^2 = 0.44$), and the range of values between individual adults (η_ω: 0.73 - 0.92) was greater than the model-difference between children and adults (Minima/Maxima of fitted curve: 0.70 -- 0.84). Taken together, these results indicate that auditory integration does not mature until around 11 years, but that the developmental effect in late childhood is small, relative to the amount of individual variability between listeners.

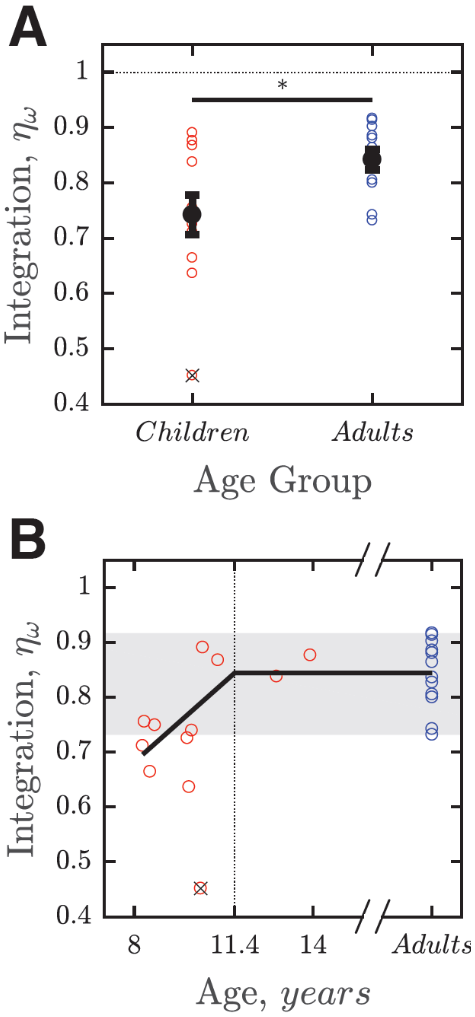


FIG 8. Experiment II: Integration efficiency for children and adults. (***A***) Group-mean [± 1 S.E.] integration efficiency (same data as Fig 6). Markers indicate values of η_ω for individual subjects (one outlier at {10.2, 0.45} was excluded from analysis, but is shown here for completeness). Horizontal dashed line represents the ideal observer. (***B***) Integration efficiency as a function of age. The solid line represents the best-fitting piecewise polynomial ('broken-stick') curve, in which the point inflection (dashed vertical line) was a free parameter. The grey shaded region indicates the 95% population interval for the adults.

C. Interim Discussion

As per Experiment I, the results of Experiment II confirmed that children are able to integrate successive observations of an auditory location cue in order to perform a perceptual averaging task, but that neither children nor adults are, on average, ideal. Unlike Experiment I, however, a significant difference was observed between children and adults, with younger children tending to be less capable integrators than adults -- only reaching adult-like performance by approximately 11 years of age.

This qualitative difference between experiments can be most parsimoniously attributed to the use of a more accurate methodology in Experiment II. Thus, as discussed after Experiment I, it is likely that at least some internal noise is irreducible, and will remain present even as N tends towards infinity. The explicit addition of reducible external noise is expected to have swamped any residual effects of irreducible internal noise, thereby providing a more accurate measure of efficiency in Experiment II.

Experiment II further allowed us to study why and in what way individual listeners were suboptimal. Typically, the pattern was towards primacy and/or recency, with listeners giving too great an importance to the first/last observation. There was, however, considerable individual variability, with many listeners exhibiting their own individual listening strategies.

The tendency of some listeners to overweight the first observation is reminiscent of the Precedence Effect, whereby multiple sounds presented in quick succession are heard as a single “fused” image whose perceived direction is skewed towards the location of the first-arriving sound (for a review, see Reference [30]). This is a primarily low-level, sensory phenomenon that ensures perceptual robustness by effectively filtering-out acoustic reflections in reverberant environments, and is subserved primarily by peripheral adaptation and inhibition in the brainstem. It is, however, unlikely to have contributed significantly to the present results for four main reasons. First, the stimulus properties are mismatched. Thus, convergent data from human psychophysics and animal physiology indicate that localization dominance occurs for lead-lag delays only up to approximately 10 ms[30]. This is an order of magnitude less than the 100 ms ISI used in the present study. And while the temporal window of the Precedence Effect has been found to increase to around 15—30 ms when stimuli are presented repeatedly[31],[32] ("buildup") --- or up to 50 ms when speech stimuli are used[33], these values still remain well-below the current ISI of 100 ms. Second, no detectable perception of fusion or echo was observed subjectively during piloting. Third, the development time-course is mismatched. For simple stimuli the Precedence Effect is believed to be adultlike by around 5 years[34],[35]. It therefore seems unable to explain the differences observed between older (8-14-year-old) children and adults in the present study. Forth and finally, the Precedence Effect primarily biases perceived direction towards the first sound (though limited up-weighting of the final sound has also been reported in some listeners[36–38]). It therefore cannot explain the substantial individual variability in weight profiles observed in the present study (see Figure 7). In short, while we cannot rule out its influence completely, the Precedence Effect seems unlikely to be a significant factor in understanding the present data. Instead the individual and developmental differences observed appear more likely due to higher-order, cognitive factors relating to perceptual decision-making (see General Discussion).

Notably, however, the Precedence Effect is itself not an entirely a low-level phenomenon, and can also be affected by various cognitive factors, including the listener's expectations (see Reference [[39]]). Some relationship with the present findings therefore cannot be ruled out altogether, and it remains an empirical question whether there is any correlation between performance on the present task, and children's ability to perceptually fuse rapid sound sequences.

IV. GENERAL DISCUSSION

The aim of this study was to quantify how integration efficiency develops during childhood. Using a multiple-observation, absolute-localization task it was shown that adults and older children are capable of integrating auditory information across sequential observations. However, the efficiency of both groups fell well below that of the ideal observer. Using Reverse Correlation, this inefficiency was shown to manifest differently across individuals, although there was a general tendency towards primacy/recency listening profiles. In terms of development, children were found to be significantly less efficient than adults, and only reached adult-like efficiency by around 11.4 years. However, the amount of development was relatively small compared to individual variability between adult listeners. Taken as a whole, the data indicates that perceptual averaging undergoes a protracted, but relatively gradual period of development during older childhood.

A. Integration efficiency in children

Among studies of audition, the present data are most comparable to those of Leibold and Bonino (2009)[15]. There, it was found that children's detection thresholds for a pure signal in noise improved progressively as the signal was repeated from 1 to 5 times. Furthermore, as in Experiment I of the present study, the rate of improvement was similar among both children and adults. These data provide converging evidence for the notion that children (in that study, as young as five years) are capable of integrating sequential auditory observations.

Outside of audition, the idea that that children are less efficient integrators is consistent with an extensive literature. For example, studies of multi-sensory integration have found young children to overly fixate on individual cues on tests of navigation[4], size/orientation discrimination[5], and stimulus detection[6]. While, in the general decision-making literature, young children have been shown to be worse at combining purely conceptual constructs, such as probabilistic information[40],[41], or risk-versus-reward[42]–[44].

It has been suggested previously that the ability to integrate sensory information only reaches maturations relatively late in a child's development[8]. In the present task, children’s behavior became adult-like at approximately 11 years. This developmental time course is in good agreement with studies of visual cue integration, where adult-like performance has been found to emerge around 11-12 years[9],[10]. However, the developmental effect in the present study was modest. It was not detectable in Experiment I, and in Experiment II the effect size was small relative to overall individual variability, with several younger children (< 11 years) performing as well as some adults. Thus, while the present data support the general notion that perceptual decision making continues to develop all throughout childhood, the changes in older childhood appear relatively small.

B. Integration efficiency in adults

The finding that adults integrate sequential information sub-optimally is consistent with several recent studies in vision. For example, Juni & Maloney (2012)[13] performed a visual analog of Experiment II. Adult observers made seven, sequential observations of a stochastic location cue (with additive jitter noise), and likewise exhibited effective, but suboptimal integration. Also as in the present study, considerable individual variability in weight vectors was observed. Thus, recency effects were particularly noticeable in some listeners, while others favored early or central intervals (see Figs A2 & A3 of Reference [[13]]). Similar findings for judgments of visual size, position, and direction have also been reported[14].

Within audition, the data from adults are also consistent with a number of previous works; in particular, a study by Swets and colleagues[12] in which listeners were asked to detect a tone presented 1 to 5 times (sequentially). As in the present study, listeners exhibited clear evidence of integration, but at a rate that was highly variable between individuals, and which generally fell markedly below that of the ideal observer[45]. Furthermore, as in the present study, integration efficiency improved markedly when external noise was added independently to each observation. This is consistent with the notion that some internal noise is non-reducible, and that this component is great enough limit the benefits of integration under noiseless listening conditions. More generally, adult performance is also consistent with a number of other 'multiple-observation' tasks such as profile analysis[26],[46] and sample discrimination[47] in audition, or motion-averaging, in vision[48], wherein it is often observed that listeners use only a fraction of the information available, and exhibit substantial individual variability in terms of which – and how many – channels they attend to.

C. Potential causes of inefficiency

Why did many individuals, and younger children in particular, fail to integrate information efficiently?

One possibility is that the observed deficits are primarily perceptual, and that information is being lost at the point of encoding due to interference --- either neural or acoustic --- between each sensory observation. In favor of this is the fact that children are also known to exhibit elevated levels of backwards-masking, and that, as in the present work, this deficit declines to near adult-levels by around 11 years[49]. Against this, however, stands the fact that sounds in the present study were separated by relatively long inter-stimulus intervals (100 ms): by which point any effects of non-simultaneous-masking are generally long-since abolished[50],[51] (see also the discussion regarding the Precedence Effect in Experiment II). Furthermore, it is difficult to see how perceptual interference could explain the level of individual variability in weight-vectors observed in Experiment II. Nor can it explain why the inefficiencies observed in adults are preserved across different tasks and sensory modalities. In short, while perceptual interference is attractive in its simplicity, it appears inconsistent with the nature of the stimuli and the pattern of data observed. This 'perceptual interference' hypothesis could be tested empirically by increasing the temporal interval or acoustic dissimilarity between observations, in which case the relative inefficiency of younger children should be diminished.

A second possibility is that inefficiencies observed in some listeners fundamentally represent limited processing capacity. Thus, a rational strategy for a system with limited memory or attention would be to fixate on a subset of the available information channels. Working memory in particular may be a limiting factor in the present study, due to the long stimulus sequence and slow presentation rate. Thus, information may have been lost over the course of the trial either due to memory decay (though cf. Fig 6B) and/or interference between the memory of each observation (see Reference [[52]]). Consistent with this, several listeners up-weighted the first/last observation: a common strategy in memory-limited tasks. Furthermore, the developmental time-course in the present study is also broadly consistent with reports that working memory continues to improves up until the age of at least 11 years old[53],[54]. This 'working memory' hypothesis predicts a correlation between efficiency in the present task, and measures of auditory working memory[55]. It also predicts that children's efficiency would progressively decrease if the memory component of the task was made more demanding (i.e., by increasing the *N* observations, or adding a second 'dual' task). Alternatively, if the number of cues were reduced, then the relative difference between children and adults should be diminished.

The idea that performance is primarily memory-limited appears plausible. However, it would be premature to assume that children's poorer performance necessarily reflects a lack of capacity. Consider, for example, a recent study in which children aged 6 to 11 years were asked to 'find the middle' of *N* simultaneously presented visual stimuli (dots). There, it was observed that children were less precise in their responses than adults: a pattern consistent with the use of only a subset of the available stimuli (i.e., due to a lack of capacity). Notably though, as the number of stimuli increased from 5 to 15, children actually became faster and more adult like in their responses. On close inspection, this change in performance appeared to be related to shift in response strategy. With small numbers of stimuli (< 6), children's trial-by-trial responses were best predicted by a strategy of 'finding the smallest shape that enclosed the visible dots, and pointing to its center' rather than the ideal strategy of computing the arithmetic mean of the individual points. The precise reason for this difference in response strategy is unknown. However, what those data demonstrate is that poor performance does not necessarily imply the inability to implement an ideal strategy efficiently. Instead, children in the present task may be opting to interpret the task in a qualitatively different way to adults (i.e., and may even be implementing a different strategy in an optimal manner). Such differences in task interpretation are difficult to evidence. However, it could be achieved, in general terms, by formulating an alternative response model that predicts an individual's trial-by-trial responses more reliably than the vector-weighted sum of the individual observations.

Fourth, a related class of explanation is that children may simply be slower to learn what the task-relevant information is, or how to weight each channel appropriately. In this respect, it is interesting to compare the present task, which requires multiple channels of useful information to be combined, with tasks of the inverse form, in which channels containing signal and noise must be segregated. For instance, studies by Kopco and colleagues have found that lateralization judgments in adults can, depending on the stimulus parameters, be biased towards or away from a preceding distractor presented at a fixed location[56],[57]. Similar, but even greater effects have also been reported in children, where, unlike in adults[56],[57], distractor-induced bias have been observed even when the perceptual similarity between target and distractor is substantial[58]. Taken together with the present study, the fact that children appear to struggle both with over-integration of useless information (in the case of distractor tasks), and under-integration of useful information (in the present study), would seem to point towards a more generalized deficit in children's ability to identify and/or attend to task relevant information. Such considerations also bring to mind Informational Masking (masking by energetically weak but unpredictable distractors), which is also elevated in young children[59], and which has likewise been attributed to an over-integration of information (this time across frequency rather than space; i.e., a broad 'attentional filter'[59],[60]). Notably, the ability to listen selectively on Informational Masking tasks has been found to improve with practice in adults[61–63]. This suggests that even for individual adults, performance on the present multiple-observation task may be limited by their ability to learn the task statistics. Furthermore, it may be that younger children are simply slower, on average, to learn the extent to which each channel contains task-relevant information. This 'slow learning' hypothesis predicts that the developmental effect would be reduced given sufficient practice, or may increase if the task-statistics were made more complex (i.e., adding different levels of external noise to each observation interval[13],[29]).

Fifth and finally, it may be that some listeners voluntarily chose not to integrate across all of the available observations. This might have happened if, for example, a listener came to suspect that some observation intervals were unreliable, or that not all observations originated from the same source location. Efforts were taken to ensure that the latter did not occur (see Experiment II Methods), and anecdotally no such suspicions were reported. It is also not immediately apparent why this would produce less integration in young children, nor why it would lead to the various patterns of weights observed in Figure 7. For instance, the most parsimonious strategy if one believed that the sounds were independent, would be to respond based on only a single observation. Such a strategy was only observed in one listener: Child 20. (NB: Alternating reliance on different individual observations could potentially have produced the more uniform weights observed in other listeners, but is inconsistent with the observed correlation between weight-efficiency and response precision.) Furthermore, such suspicions are unlikely to explain the suboptimal integration observed Experiment I, where all observations were in fact located identically (although, due to internal noise, even identical stimuli are sometimes liable to be perceived as different[64]). Nonetheless, the possibility that some listeners chose to discount certain observations cannot be ruled out. This possibility could be investigated experimentally by systematically increasing the amount of external noise (i.e., the sigma parameter of the jitter distribution). In this case one would predict to see discontinuities, with a rapid reduction in weight-efficiency at the point where listeners started to notice discrepancies.

Listeners might also have decided to voluntarily ignore some channels for the sake of ease, assuming that the integration of each additional observation incurs some non-trivial ‘cost’ in terms of listening effort. Such differences in motivation are always a concern in developmental studies, and pains were taken to ensure that children remained engaged and focused throughout the experiment. Furthermore, from a developmental perspective, the fact that the one child (Child 20) who exhibited a relatively simple ‘single observation’ strategy was such a marked outlier in terms of efficiency is encouraging, as it suggests that younger children were not simply the ‘tail end’ of some normal distribution of motivation (see Fig 8B). However, the possibility that differences in motivation affected performance of some individuals cannot be ruled out. It could be probed empirically by including a subset of ‘high value’ trials (i.e., with an association financial incentive, or some child-friendly equivalent). If differences in motivation/effort do affect performance, then the difference between children and adults, or between individual adults, should be diminished on such trials.

D. Absolute sound localization performance in children and adults

Although the present study was concerned primarily with integration efficiency, it may also be of interest to consider how listeners’ sound-localization performance compared with data reported previously.

For adults, the present data are most comparable to the ‘noise’ condition of Recanzone, Makhamra, & Guard (1998)[65], who measured absolute-localization performance using 200 ms white noise bursts. Within the central ±17° (i.e., the range of the present study), response errors were relatively stable, with a standard deviation of approximately 5°. This is in good agreement with the present data in Experiment 1, where the group-mean standard deviation (‘imprecision’) was 4.81° for adults and 5.53 for children° (Figure 2, $N = 1$ condition). The present values are also comparable to those of Yost and Zhong (2014)[18], who asked listeners to localize 200 ms noise bursts of variable bandwidth and central frequency. There, RMS error (which, for an unbiased listener, is equivalent to the standard deviation of errors) was approximately 7.5° for a 1 octave bandpass noise centered on 2 kHz. This is somewhat higher than the value of 4.81° observed in the present study. However, it also includes presentations of up to +75°, and localization ability is known to decrease with eccentricity[18]. Conversely, at a single eccentricity of +15°, Yost and Zhong reported a mean RMS error of approximately 4° for bandwidths between 1/6 to 2 octaves: a value that is roughly consistent with the present value of 4.81° (measured with a bandwidth of 1 octave only).

For children, we are aware of no directly comparable data. However, the finding that children’s response precision in the N=1 condition was not significantly lower than adults is consistent with a number of studies showing that Minimal Audible Angles are largely adult-like by 5 years[34], and that absolute localization performance is mature by around 6 years[66],[67] (for a review, see Reference [3]). In short, in terms of absolute localization ability, the results of both children and adults appear to be in good agreement with previous data.

V. CONCLUSIONS

(i) Using a multiple-observation localization task, both children and adults were shown to be effective integrators: able to combine up to five sequentially presented auditory stimuli.

(ii) However, while localization precision improved as a function of *N* observations, the rate of gain was substantially less than that predicted by an ideal observer (Experiment I). This may indicate suboptimal integration. Alternatively, it may be that performance is limited by a substantial component of irreducible noise (e.g., correlated sensory noise, or response errors).

(iii) When using Reverse Correlation (Experiment II), children were shown to be less efficient integrators than adults, only exhibiting adult-like performance by ~11 years old. The developmental effect was small, however, relative to the amount of individual variability, with younger children often exhibiting greater integration efficiency than some adults. That sensory integration does not develop until around 11 years is consistent with previous studies in vision. However, the modest effect size indicates a protracted, but relatively gradual period of development during older childhood.

(iv) Substantial individual variability in listening strategy was observed. There was a general trend towards overweighting the first (primacy) or last (recency) observation. However, other patterns were also observed. The causes of the individual and developmental differences in integration efficiency remain unclear. However, five possible explanations are discussed, and testable predictions for each are detailed.

Acknowledgments

The authors would like to thank Nazia Khan and Sanggeeta Surandran for piloting and data collection, and to Dr Sara Garcia and the UCL Institute of Ophthalmology Workshop for their assistance in the design and construction of the speaker ring. This work was supported by the NIHR Biomedical Research Centre located at (both) Moorfields Eye Hospital and the UCL Institute of Ophthalmology.

REFERENCES

[1]L. A. Werner and E. W. Rubel, *Developmental Psychoacoustics.* (American Psychological Association, 1992).

[2]N. C. Thompson, J. L. Cranford, and E. Hoyer, "Brief-tone frequency discrimination by children," J. Speech, Lang. Hear. Res. **42**, 1061–1068 (1999).

[3]C. Freigang, N. Richter, R. Rübsamen, and A. A. Ludwig, "Age-related changes in sound localisation ability," Cell Tissue Res. **361**, 371–386 (2015).

[4]M. Nardini, P. Jones, R. Bedford, and O. Braddick, "Development of cue integration in human navigation," Curr. Biol. **18**, 689–693 (2008).

[5]M. Gori, M. Del Viva, G. Sandini, and D. C. Burr, "Young children do not integrate visual and haptic form information," Curr. Biol. **18**, 694–698 (2008).

[6]A. B. Brandwein, J. J. Foxe, N. N. Russo, T. S. Altschuler, H. Gomes, and S. Molholm, "The development of audiovisual multisensory integration across childhood and early adolescence: a high-density electrical mapping study," Cereb. Cortex **21**, 1042–1055 (2011).

[7]D. Burr and M. Gori, "Multisensory integration develops late in humans," in *The Neural Bases of Multisensory Processes*, M. M. Murray and M. T. Wallace, eds. (CRC Press/Taylor & Francis, 2012).

[8]M. O. Ernst, "Multisensory integration: a late bloomer," Curr. Biol. **18**, R519--R521 (2008).

[9]M. Nardini, R. Bedford, and D. Mareschal, "Fusion of visual cues is not mandatory in children," Proc. Natl. Acad. Sci. U. S. A. **107**, 17041–17046 (2010).

[10]T. M. Dekker, H. Ban, B. van der Velde, M. I. Sereno, A. E. Welchman, and M. Nardini, "Late development of cue integration is linked to sensory fusion in cortex," Curr. Biol. **25**, 2856–2861 (2015).

[11]P. Allen, R. Jones, and P. Slaney, "The role of level, spectral, and temporal cues in children's detection of masked signals," J. Acoust. Soc. Am. **104**, 2997–3005 (1998).

[12]J. A. Swets, E. F. Shipley, M. J. McKey, and D. M. Green, "Multiple Observations of Signals in Noise," J. Acoust. Soc. Am. **31**, 514–521 (1959).

[13]M. Z. Juni, T. M. Gureckis, and L. T. Maloney, "Effective integration of serially presented stochastic cues," J. Vis. **12**, 1–16 (2012).

[14]B. Hubert-Wallander and G. M. Boynton, "Not all summary statistics are made equal: Evidence from extracting summaries across time," J. Vis. **15**, 1–12 (2015).

[15]L. J. Leibold and A. Y. Bonino, "Release from informational masking in children: Effect of multiple signal bursts," J. Acoust. Soc. Am. **125**, 2200–2208 (2009).

[16]P. R. Jones, "A tutorial on cue combination and Signal Detection Theory: Using changes in sensitivity to evaluate how observers integrate sensory information," J. Math. Psychol. **73**, 117–139 (2016).

[17]"See supplementary material at [URL will be inserted by BJ] for additional supporting material," (n.d.).

[18]W. A. Yost and X. Zhong, "Sound source localization identification accuracy: Bandwidth dependencies," J. Acoust. Soc. Am. **136**, 2737–2746 (2014).

[19]D. Kostadinov and J. D. Reiss, "Spatial Audio Matlab Toolbox," (2009).

[20]D. H. Brainard, "The psychophysics toolbox," Spat. Vis. **10**, 433–436 (1997).

[21]P. R. Jones, S. E. Garcia, and M. Nardini, "Digital LED Pixels: Instructions for use and a characterization of their properties," Behav. Res. Methods **48**, 1266–1284 (2016).

[22]H. Boyaci, K. Doerschner, and L. T. Maloney, "Cues to an equivalent lighting model," J. Vis. **6**, 2 (2006).

[23]M. S. Landy, M. S. Banks, and D. C. Knill, "Ideal-observer models of cue integration," in *Sensory Cue Integration*, J. Trommershauser, K. Kording, and M. S. Landy, eds. (Oxford University Press, 2011).

[24]S. He, E. Buss, and J. W. Hall III, "Monaural temporal integration and temporally selective listening in children and adults," J. Acoust. Soc. Am. **127**, 3643–3653 (2010).

[25]D. G. Pelli, "Noise in the visual system may be early," in *Computational Models of Visual Processing*, M. L. & J. A. Movshon, ed. (MIT Press, 1991), pp. 147–152.

[26]B. G. Berg, "A molecular description of profile analysis: Decision weights and internal noise," J. Acoust. Soc. Am. **115**, 822–829 (2004).

[27]R. A. Lutfi, "Correlation coefficients and correlation ratios as estimates of observer weights in multiple-observation tasks," J. Acoust. Soc. Am. **97**, 1333–1334 (1995).

[28]V. M. Richards and S. Zhu, "Relative estimates of combination weights, decision criteria, and internal noise based on correlation coefficients," J. Acoust. Soc. Am. **95**, 423–434 (1994).

[29]H. Dai and B. G. Berg, "Spectral and temporal weights in spectral-shape discrimination," J. Acoust. Soc. Am. **92**, 1346–1355 (1992).

[30]A. D. Brown, G. C. Stecker, and D. J. Tollin, "The precedence effect in sound localization," J. Assoc. Res. Otolaryngol. **16**, 1–28 (2015).

[31]T. Djelani and J. Blauert, "Investigations into the build-up and breakdown of the precedence effect," Acta Acust. united with Acust. **87**, 253–261 (2001).

[32]R. L. Freyman, R. K. Clifton, and R. Y. Litovsky, "Dynamic processes in the precedence effect," J. Acoust. Soc. Am. **90**, 874–884 (1991).

[33]J. P. A. Lochner and J. F. Burger, "The subjective masking of short time delayed echoes by their primary sounds and their contribution to the intelligibility of speech," Acta Acust. united with Acust. **8**, 1–10 (1958).

[34]R. Y. Litovsky, "Developmental changes in the precedence effect: estimates of minimum audible angle," J. Acoust. Soc. Am. **102**, 1739–1745 (1997).

[35]R. Y. Litovsky, H. S. Colburn, W. A. Yost, and S. J. Guzman, "The precedence effect," J. Acoust. Soc. Am. **106**, 1633–1654 (1999).

[36]G. C. Stecker, J. D. Ostreicher, and A. D. Brown, "Temporal weighting functions for interaural time and level differences. III. Temporal weighting for lateral position judgments," J. Acoust. Soc. Am. **134**, 1242–1252 (2013).

[37]G. C. Stecker, "Temporal weighting functions for interaural time and level differences. IV. Effects of carrier frequency," J. Acoust. Soc. Am. **136**, 3221–3232 (2014).

[38]G. C. Stecker and E. R. Hafter, "Temporal weighting in sound localization," J. Acoust. Soc. Am. **112**, 1046–1057 (2002).

[39]R. K. Clifton and R. L. Freyman, "The precedence effect: Beyond echo suppression," in *Binaural and Spatial Hearing in Real and Virtual Environments* (New York: Erlbaum, 1997), pp. 233–256.

[40]T. Betsch and A. Lang, "Utilization of probabilistic cues in the presence of irrelevant information: A comparison of risky choice in children and adults," J. Exp. Child Psychol. **115**, 108–125 (2013).

[41]T. Betsch, A. Lehmann, S. Lindow, A. Lang, and M. Schoemann, "Lost in search: (Mal--) adaptation to probabilistic decision environments in children and adults.," Dev. Psychol. **52**, 311–325 (2016).

[42]A. Schlottmann, "Children's judgements of gambles: A disordinal violation of utility," J. Behav. Decis. Mak. **13**, 77–89 (2000).

[43]T. W. Boyer, "The development of risk-taking: A multi-perspective review," Dev. Rev. **26**, 291–345 (2006).

[44]J. A. Weller, I. P. Levin, and N. L. Denburg, "Trajectory of risky decision making for potential gains and losses from ages 5 to 85," J. Behav. Decis. Mak. **24**, 331–344 (2011).

[45] and is not directly comparable to the parameter k in the present paper; though both predict an improvement of √N when integration is maximal. W. the data in T. I. 2 of S. et al (1959) are refit using E. 2 then k for the three reported observer {NB: the parameter k in Swets et al (1959) extends from 0 to infinity, "ENDNOTE," (n.d.).

[46]D. M. Green and C. R. Mason, "Auditory profile analysis: Frequency, phase, and Weber's Law," J. Acoust. Soc. Am. **77**, 1155–1161 (1985).

[47]B. G. Berg, "Analysis of weights in multiple observation tasks," J. Acoust. Soc. Am. **86**, 1743–1746 (1989).

[48]S. C. Dakin, "Information limit on the spatial integration of local orientation signals," JOSA A **18**, 1016–1026 (2001).
[49]D. E. H. Hartley, B. A. Wright, S. C. Hogan, and D. R. Moore, "Age-related improvements in auditory backward and simultaneous masking in 6-to 10-year-old children," J. Speech, Lang. Hear. Res. **43**, 1402–1415 (2000).
[50]A. J. Oxenham and B. C. J. Moore, "Modeling the additivity of nonsimultaneous masking," Hear. Res. **80**, 105–118 (1994).
[51]N. Weisstein and R. N. Haber, "A U-shaped backward masking function in vision," Psychon. Sci. **2**, 75–76 (1965).
[52]J. S. Snyder and D. M. Weintraub, "Loss and persistence of implicit memory for sound: Evidence from auditory stream segregation context effects," Attention, Perception, Psychophys. **75**, 1059–1074 (2013).
[53]J. Johnson, V. Fabian, and J. Pascual-Leone, "Quantitative hardware stages that constrain language development," Hum. Dev. **32**, 245–271 (1989).
[54]L. Henry, *The Development of Working Memory in Children* (Sage, 2011).
[55]L. S. Siegel and E. B. Ryan, "The development of working memory in normally achieving and subtypes of learning disabled children," Child Dev. **60**, 973–980 (1989).
[56]L. Hládek, B. Tomoriová, and N. Kopčo, "Temporal characteristics of contextual effects in sound localization," J. Acoust. Soc. Am. **142**, 3288–3296 (2017).
[57]N. Kopčo, V. Best, and B. G. Shinn-Cunningham, "Sound localization with a preceding distractor," J. Acoust. Soc. Am. **121**, 420–432 (2007).
[58]K. Petrini, P. R. Jones, L. Smith, and M. Nardini, "Hearing Where the Eyes See: Children Use an Irrelevant Visual Cue When Localizing Sounds," Child Dev. **86**, 1449–1457 (2015).
[59]F. L. Wightman, M. R. Callahan, R. A. Lutfi, D. J. Kistler, and E. Oh, "Children’s detection of pure-tone signals: Informational masking with contralateral maskers," J. Acoust. Soc. Am. **113**, 3297–3305 (2003).
[60]P. R. Jones, D. R. Moore, and S. Amitay, "Development of Auditory Selective Attention : Why Children Struggle to Hear in Noisy Environments," Dev. Psychol. **51**, 353–369 (2015).
[61]P. R. Jones, D. R. Moore, D. E. Shub, and S. Amitay, "Learning to detect a tone in unpredictable noise," J. Acoust. Soc. Am. **135**, EL128--EL133 (2014).
[62]J. M. Gold, A. B. Sekuler, and P. J. Bennett, "Characterizing perceptual learning with external noise," Cogn. Sci. **28**, 167–207 (2004).
[63]R. W. Li, D. M. Levi, and S. A. Klein, "Perceptual learning improves efficiency by re-tuning the decision’template’for position discrimination," Nat. Neurosci. **7**, 178–183 (2004).
[64]S. Amitay, D. R. Moore, K. Molloy, and L. F. Halliday, "Feedback valence affects auditory perceptual learning independently of feedback probability," PLoS One **10**, e0126412 (2015).
[65]G. H. Recanzone, S. D. D. R. Makhamra, and D. C. Guard, "Comparison of relative and absolute sound localization ability in humans," J. Acoust. Soc. Am. **103**, 1085–1097 (1998).
[66]R. J. Otte, M. J. H. Agterberg, M. M. Van Wanrooij, A. F. M. Snik, and A. J. Van Opstal, "Age-related hearing loss and ear morphology affect vertical but not horizontal sound-localization performance," J. Assoc. Res. Otolaryngol. **14**, 261–273 (2013).
[67]S. Kühnle, A. A. Ludwig, S. Meuret, C. Küttner, C. Witte, J. Scholbach, M. Fuchs, and R. Rübsamen, "Development of auditory localization accuracy and auditory spatial discrimination in children and adolescents," Audiol. Neurotol. **18**, 48–62 (2012).