Optimal Audiovisual Integration in the Ventriloquism Effect But Pervasive Deficits in Unisensory Spatial Localization in Amblyopia

Michael D. Richards,1–3 Herbert C. Goltz,3,4 and Agnes M. F. Wong2–4

1Institute of Medical Science, Faculty of Medicine, University of Toronto, Toronto, Canada
2Department of Ophthalmology and Vision Sciences, The Hospital for Sick Children, Toronto, Canada
3Department of Ophthalmology and Vision Sciences, University of Toronto, Toronto, Canada
4Program in Neurosciences and Mental Health, The Hospital for Sick Children, Toronto, Canada

Purpose. Classically understood as a deficit in spatial vision, amblyopia is increasingly recognized to also impair audiovisual multisensory processing. Studies to date, however, have not determined whether the audiovisual abnormalities reflect a failure of multisensory integration, or an optimal strategy in the face of unisensory impairment. We use the ventriloquism effect and the maximum-likelihood estimation (MLE) model of optimal integration to investigate integration of audiovisual spatial information in amblyopia.

Methods. Participants with unilateral amblyopia (n = 14; mean age 28.8 years; 7 anisometropic, 3 strabismic, 4 mixed mechanism) and visually normal controls (n = 16, mean age 29.2 years) localized brief unimodal auditory, unimodal visual, and bimodal (audiovisual) stimuli during binocular viewing using a location discrimination task. A subset of bimodal trials involved the ventriloquism effect, an illusion in which auditory and visual stimuli originating from different locations are perceived as originating from a single location. Localization precision and bias were determined by psychometric curve fitting, and the observed parameters were compared with predictions from the MLE model.

Results. Spatial localization precision was significantly reduced in the amblyopia group compared with the control group for unimodal visual, unimodal auditory, and bimodal stimuli. Analyses of localization precision and bias for bimodal stimuli showed no significant deviations from the MLE model in either the amblyopia group or the control group.

Conclusions. Despite pervasive deficits in localization precision for visual, auditory, and audiovisual stimuli, audiovisual integration remains intact and optimal in unilateral amblyopia.

Keywords: amblyopia, multisensory integration, audiovisual integration, ventriloquism effect, spatial localization

Amblyopia is a unilateral, or rarely bilateral, developmental visual disorder that affects 2% to 4% of the population.1 Its associated visual impairment cannot be directly attributed to a structural eye abnormality, but arises from amblyogenic factors that interfere with normal visual experience during a sensitive period in early life. The most common amblyogenic factors are anisometropia (difference in refractive error between the eyes), strabismus (eye misalignment), or a combination of both.2 More rarely, amblyopia may also be caused by visual deprivation from early-onset cataract or ptosis. In addition to the reduction in visual acuity by which it is clinically defined,3 amblyopia is associated with a range of visual deficits including impairments in contrast sensitivity, stereo acuity, recognition of ‘crowded’ optotypes, global motion detection, and real-world scene perception.4–8 The disorder also involves spatial distortions and positional uncertainty that affect vision in both the amblyopic eye and the fellow eye.9–15 Emerging research suggests that perceptual abnormalities in amblyopia are not limited to vision, but also involve multisensory integration in the audiovisual realm.14–16 Multisensory integration refers to the process of unifying overlapping information from different senses to form an accurate and coherent perception of the external world.17 This ability is not of trivial consequence, but confers numerous adaptive advantages in terms of reaction times,18 detection sensitivity,19 discrimination reliability,20 and speech comprehension.21–24 A well-known paradigm for studying multisensory integration is the McGurk effect—a robust audiovisual illusion in which incongruity between the linguistic content of an auditory speech signal and the accompanying visual speech signal alters the perceived auditory signal.25 For example, if audio of the syllable /ba/ is paired with video of the syllable /ga/, the perceived auditory percept is typically an illusory /da/. This phenomenon generalizes to many syllabic combinations, with the resulting auditory percept being either a fused syllable intermediate to the veridical auditory and visual cues, or dominated by the visual cue.26,27 Computational modeling within a Bayesian framework suggests that the illusory McGurk percept may represent a statistically optimal strategy to integrate visual and auditory speech cues.28
Several recent studies report that susceptibility to the McGurk effect is reduced in people with unilateral amblyopia, even when viewing with both eyes.

Although this finding may indicate a failure to optimally integrate the available unimodal (e.g., visual and auditory) information, it may also alternatively result from deficiencies at the unisensory level propagated through an intact system for optimal multisensory integration. For example, if visual speech information from both eyes is not reliably encoded by the amblyopic brain, then a reduced McGurk effect (i.e., a shift toward veridical auditory perception) may reflect an optimal strategy of sensory re-weighting. However, the studies to date were not designed, nor can their results be modeled, to distinguish integration failure from sensory re-weighting. Studies of other multisensory abnormalities in amblyopia have been similarly unable to make this mechanistic distinction. For example, the ability to detect asynchrony between auditory and visual stimuli is diminished in amblyopia. While insensitivity to audiovisual asynchrony may indicate abnormal audiovisual integration, it may also be caused by noisy encoding of visual temporal information that feeds into a normal upstream multisensory processing network. Considering these possibilities, it becomes clear that the origin of these multisensory abnormalities in amblyopia remains an unresolved question. Does anomalous audiovisual perception in amblyopia result from failure to integrate the available unisensory information, or does it represent appropriate (i.e., statistically optimal) integration of the available, but deficient, unisensory information?

Unlike previous investigations of multisensory integration in amblyopia, this study aims to differentiate between deficits at the unisensory and multisensory levels of sensory processing by explicitly modeling audiovisual integration in a ventriloquism effect paradigm. The ventriloquism effect is an audiovisual illusion in which spatially disparate visual and auditory stimuli are perceived as originating from the same location. Typically, the location information of the unimodal visual component dominates in the perceived location of the fused audiovisual percept, a process sometimes termed visual capture. In a manipulation of the ventriloquism effect, a reduced McGurk effect (i.e., a shift toward veridical auditory perception) may reflect an optimal strategy of sensory re-weighting. However, the studies to date were not designed, nor can their results be modeled, to distinguish integration failure from sensory re-weighting. Studies of other multisensory abnormalities in amblyopia have been similarly unable to make this mechanistic distinction. For example, the ability to detect asynchrony between auditory and visual stimuli is diminished in amblyopia. While insensitivity to audiovisual asynchrony may indicate abnormal audiovisual integration, it may also be caused by noisy encoding of visual temporal information that feeds into a normal upstream multisensory processing network. Considering these possibilities, it becomes clear that the origin of these multisensory abnormalities in amblyopia remains an unresolved question. Does anomalous audiovisual perception in amblyopia result from failure to integrate the available unisensory information, or does it represent appropriate (i.e., statistically optimal) integration of the available, but deficient, unisensory information?

Audiovisual spatial integration and the ventriloquism effect offer several advantages over the McGurk effect as an approach for studying audiovisual integration in amblyopia. First, it involves simple spatial signals that are measured on a continuous scale, as opposed to complex linguistic signals that are measured categorically. Categoric syllabic data are not only more difficult to transcribe faithfully, but also more challenging to model mathematically. Second, the ventriloquism effect is not significantly affected by top-down directed attention or bottom-up automatic attention. In contrast, measures of integration using the McGurk effect may be confounded by attentional effects, as it is significantly diminished by increased attentional load. Finally, multisensory enhancement in spatial localization precision, if observed in amblyopia, will provide clear evidence of audiovisual integration, rather than nonintegrative unisensory dominance or perceptual blending.

The MLE model of audiovisual spatial integration offers a powerful methodology to distinguish two possible causes of the multisensory abnormalities observed in amblyopia: optimal integration of deficient unisensory information, or failure to integrate the available unisensory information. If integration is intact, then predictions of the MLE model will be upheld despite deficits at the unisensory level. Specifically, bimodal localization will show precision enhancement over unimodal localization consistent with the MLE model, and spatial bias in the ventriloquism effect will be consistent with MLE-optimal combination of the unimodal position signals. Conversely, if integration is impaired, then spatial precision of audiovisual localization will show less enhancement than predicted by the MLE model, and spatial bias in the ventriloquism effect will not follow the MLE model.

**METHODS**

**Participants**

Two groups of participants were recruited: adults with unilateral amblyopia, and adults with normal vision. Participants were excluded if they had manifest or latent nystagmus, or a history of neurodevelopmental or neurological disorder, hearing impairment, high ametropia (hyperopia > +5 diopters [D] or myopia > −6D), or any other ocular pathology or prior intraocular surgery. Each participant underwent an ocular and hearing screen by a certified orthoptist or ophthalmologist. The ocular assessment measured habitual refractive correction (automatic lensmeter), distance visual acuity (Early Treatment Diabetic Retinopathy Study chart with habitual correction), stereo acuity (Randot circles and Titmus fly test), foveal suppression (Worth 4-dot test), and eye alignment (prism cover test). The hearing assessment ensured reliable detection of suprathreshold pure tones (25-dBA sound pressure level) in each ear at four standard frequencies (500, 1000, 2000, and 4000 Hz) using a screening audiometer (model MA 27; MAICO Diagnostics, Eden Prairie, MN, USA) with circumaural headphones (model TDH 39; MAICO Diagnostics). Amblyopia was defined as visual acuity of 0.18 logMAR (20/30) or poorer in the amblyopic eye, visual acuity of at least 0.1 logMAR (20/25) in the fellow eye, and an interocular acuity difference of 0.2 logMAR or more. Participants were classified as having anisometropic amblyopia if the interocular difference in spherical equivalent or cylindrical error was 1 D or more, as having strabismic amblyopia if there was any manifest deviation on cover testing in the absence of anisometropia, or as having mixed-mechanism amblyopia when both anisometropia and a manifest deviation of ≥8 prism diopters or more were present. Visually normal was defined as visual acuity of at least 0.1 logMAR (20/25) in each eye, with stereo acuity of 40 seconds of arc or better, and no manifest strabismus. The Research Ethics Board at The Hospital for Sick Children approved the study, and all protocols adhered to the tenets of the Declaration of Helsinki. Written informed consent was obtained from all participants after explanation of the nature and possible consequences of the study.

Fourteen adults with unilateral amblyopia (mean age ± SD range: 28.8 ± 8.9 [19–48] years; subtypes: 7 anisometropic, 3 strabismic, 4 mixed) and 16 visually normal controls (mean age ± SD range: 29.2 ± 7.1 [23–47] years) participated in the study. Individuals were not age-matched, but the two groups did not differ significantly in age (t28 = 0.137, P = 0.892).

Clinical characteristics of the participants with amblyopia are summarized in Table 1. The sample size provided 80% power to detect an effect size of d = 0.81 for within-subjects effects in the amblyopia group (e.g., comparing observed versus MLE-predicted localization precision), and an effect size of d = 1.06 for between-subjects effects (e.g., comparing localization
Table 1. Characteristics of Participants with Amblyopia

<table>
<thead>
<tr>
<th>ID</th>
<th>Age (y)</th>
<th>Subtype</th>
<th>Visual Acuity (LogMAR)</th>
<th>Refractive Correction</th>
<th>Stereo Acuity (arc sec)</th>
<th>Worth 4-Dot Response</th>
<th>Additional Details</th>
</tr>
</thead>
<tbody>
<tr>
<td>A1</td>
<td>29</td>
<td>Strab</td>
<td>0.00 1.00</td>
<td>None</td>
<td>None</td>
<td>Negative</td>
<td>LE suppress</td>
</tr>
<tr>
<td>A2</td>
<td>22</td>
<td>Aniso</td>
<td>0.00 0.48</td>
<td>+1.50 +0.50 × 80</td>
<td>+1.00 +1.25 × 95</td>
<td>200</td>
<td>Fused</td>
</tr>
<tr>
<td>A3</td>
<td>48</td>
<td>Aniso</td>
<td>0.70 0.00</td>
<td>+2.25 +0.25 × 174</td>
<td>-1.00 +1.25</td>
<td>3000</td>
<td>Fused</td>
</tr>
<tr>
<td>A4</td>
<td>29</td>
<td>Aniso</td>
<td>0.48 -0.10</td>
<td>-5.00</td>
<td>+1.25</td>
<td>3000</td>
<td>Fused</td>
</tr>
<tr>
<td>A5</td>
<td>23</td>
<td>Aniso</td>
<td>-0.10 0.48</td>
<td>-2.25</td>
<td>+0.25 +2.25 × 85</td>
<td>200</td>
<td>Fused</td>
</tr>
<tr>
<td>A6</td>
<td>29</td>
<td>Mixed</td>
<td>0.00 1.00</td>
<td>Plano</td>
<td>+3.50 +2.00 × 90</td>
<td>40</td>
<td>Fused</td>
</tr>
<tr>
<td>A7</td>
<td>19</td>
<td>Aniso</td>
<td>0.00 0.18</td>
<td>-0.75 +2.00 × 84</td>
<td>-0.75 +1.25 × 99</td>
<td>200</td>
<td>Fused</td>
</tr>
<tr>
<td>A8</td>
<td>27</td>
<td>Strab</td>
<td>0.00 0.48</td>
<td>+3.25 +1.00 × 43</td>
<td>-0.25 +1.25 × 135</td>
<td>40</td>
<td>Fused</td>
</tr>
<tr>
<td>A9</td>
<td>37</td>
<td>Mixed</td>
<td>-0.10 1.30</td>
<td>-1.00</td>
<td>+0.25 +2.50 × 120</td>
<td>140</td>
<td>Fused</td>
</tr>
<tr>
<td>A10</td>
<td>32</td>
<td>Aniso</td>
<td>-0.10 0.54</td>
<td>Plano</td>
<td>+2.00 +2.00 × 124</td>
<td>350</td>
<td>Fused</td>
</tr>
<tr>
<td>A11</td>
<td>23</td>
<td>Strab</td>
<td>0.20 0.00</td>
<td>+0.50 +0.50 × 28</td>
<td>+1.25 +0.50 × 88</td>
<td>350</td>
<td>Fused</td>
</tr>
<tr>
<td>A12</td>
<td>44</td>
<td>Mixed</td>
<td>0.90 0.00</td>
<td>+6.00 +1.25 × 75</td>
<td>-0.75</td>
<td>Negative</td>
<td>RE suppress</td>
</tr>
<tr>
<td>A13</td>
<td>22</td>
<td>Aniso</td>
<td>1.10 -0.10</td>
<td>-6.00 +0.75 × 174</td>
<td>-4.50 +0.50 × 75</td>
<td>350</td>
<td>Fused</td>
</tr>
<tr>
<td>A14</td>
<td>19</td>
<td>Mixed</td>
<td>0.48 0.00</td>
<td>+3.00 +1.00 × 130</td>
<td>+4.25</td>
<td>350</td>
<td>Fused</td>
</tr>
</tbody>
</table>

RE, right eye; LE, left eye; Aniso, anisometropia; Strab, strabismic.

precision between groups), for a two-tailed test at alpha = 0.05 (computed using G*Power 3.1).14

Apparatus and Stimuli

The entire experiment was conducted in a darkened, double-walled audiometric chamber (internal dimensions 2.0 × 2.1 × 2.2 m). The floor was carpeted, and the walls and ceiling were lined with 5-cm acoustic wedge foam (Foam Factory, Macomb, MI, USA). Head position was constrained by a chinrest fixed to a small table 65 cm from the center of the audiovisual apparatus (see Supplementary Fig. S1). Visual stimuli consisted of medium contrast (39%) Gaussian blobs of five sizes (1 SD = 16°, 20°, 24°, 28°, or 32°), flashed on a large light-emitting diode monitor (model E654; NEC Corporation, Tokyo, Japan) for 32 ms (2 frames at 60 Hz). The monitor subtended 96° × 64° of visual angle (165-cm diagonal) and was overlaid with a three-stop neutral density filter (ND 0.9) to create a Gaussian blob (peak luminance 2.1 cd/m²) with imperceptible steps between gray levels. Auditory stimuli consisted of 32-ms clicks (8 cycles of 2–5 kHz bandpass filtered white noise, 4 ms in duration, enveloped with a 2-ms sigmoid on/off ramp), presented at 62.0 dB through two speakers mounted on either side of the monitor at the horizontal midline. Apparent click location was controlled by linear amplitude panning of interaural level difference cues.42,43 The output profile for each speaker was measured across the entire stimulus dynamic range using a sound level meter to ensure their outputs were identical. Auditory and visual stimulus timings were confirmed with an oscilloscope. A wireless gamepad (model F710; Logitech, Newark, CA, USA) was used to initiate trials and enter responses.

Procedure

All trials were conducted with both eyes open. Participants performed a relative spatial localization task for unimodal stimuli (visual blobs only or auditory clicks only) and bimodal stimuli (blobs and clicks together) similar to that described by Alais and Burr.35 A general trial timeline is illustrated in Figure 1. Upon initiation of each trial, a red fixation dot (0.66°) was presented centrally for 500 ms, followed by a randomized delay of 250 to 400 ms. Two matching stimuli (a test stimulus and probe stimulus, but in random order) were then presented in succession, 500-ms apart, and the participant was asked to indicate whether the second event occurred left or right relative to the first event.” Participants were instructed to keep their head and eyes aligned centrally. There were 21 test stimulus conditions: six unimodal (1 click-only and 5 blob size variants), five bimodal with spatially congruent clicks and blobs (5 blob size variants paired with a click), and 10 bimodal with spatially conflicting clicks and blobs (5 blob size variants paired with a click, but blob displaced 4° left and click displaced 4° right, or click displaced 4° left and blob displaced 4° right). Bimodal test stimulus conditions with spatial conflict were designed to elicit the ventriloquism effect, and participants were not told of this spatial disparity. The test stimulus was presented centrally (0°) in all trials; for bimodal test stimuli with spatial conflict, the unimodal components were displaced 4° in opposite directions such that that their average location

![Figure 1](https://example.com/figure1.png)

**Figure 1.** Illustration of the trial timeline. After trial initiation by the participant, a fixation dot appeared centrally for 500 ms, followed by a dark interval of 250 to 400 ms. Two brief stimuli (test and probe) were displayed in sequence, 500-ms apart, but in random order. The participant judged whether the second stimulus originated left or right relative to the first.
was still 0°. The probe stimulus matched the characteristics of the test stimulus except for horizontal displacement (specified in Table 2), and in some cases, spatial congruency (bimodal test stimuli with spatial conflict were paired with spatially congruent probe stimuli). Data were collected in separate blocks for the unimodal auditory conditions, and for each of the five blob sizes within the unimodal visual and bimodal conditions. Twenty trials were run for each probe stimulus displacement, randomly interleaved within each block.

### Data Analysis

Psychometric curve fitting and parameter calculation were performed individually for each participant. For each test stimulus condition, the proportion of trials in which the probe was perceived ‘left’ of the test stimulus was plotted as a function of probe displacement. These data were fitted with cumulative normal functions free to vary in horizontal position (i.e., unconstrained mean and variance). Fitting was computed by the maximum likelihood method using custom-written scripts in MATLAB version 2011b (Mathworks, Inc., Natick, MA, USA). The mean of the fitted function, or point of subjective equality (PSE), represents the bias in the localization estimate of the test stimulus (Δ, Equation 1 in Appendix A). The maximum slope of the function (β′, Equation 5 in Appendix A) represents the precision of the localization estimate, and was computed from the SD of the fitted function.

In the 16°/SD unimodal visual condition, undersampling resulted in falsely steep fits for six participants (38%) in the control group and six participants (43%) in the amblyopia group. Data from 16°/SD conditions for all participants were therefore excluded from analysis.

The MLE-predicted values for bimodal localization precision (i.e., predicted β′) were calculated from the variances of the fitted functions in each unimodal condition and the specified locations of the auditory and visual signals (using Equations 1, 2, and 3 in Appendix A). All statistical tests were performed using IBM SPSS Statistics version 22 (Armonk, NY, USA). Statistical significance was defined as P < 0.05. Localization bias data from the two conflict conditions across each unimodal condition and the specified locations of the auditory and visual signals were compared by the Bonferroni method.

### Results

Representative data and fitted functions for the unimodal and bimodal localization tasks are shown in Figure 2. Subsequent analyses of localization precision, localization bias, and agreement with the MLE model are reported in the sections below.

### Localization Precision

#### Unimodal Stimuli

Localization precision, defined as the slope of the fitted psychometric function at the midpoint, decreased in both groups for unimodal visual stimuli as the blob size increased from 20°/SD to 32°/SD (Fig. 3A). A two (group) × four (blob size) mixed-design ANOVA for unimodal visual localization precision yielded a significant main effect of group (F1,28 = 4.692, P = 0.039, η̂2p = .144), a significant main effect of blob size (F2,44.68.32 = 175.45, P < 0.001, η̂2p = .862, Greenhouse-Geisser correction), but no significant interaction of group and blob size (F2,44.68.32 = 2.31, P = 0.096, η̂2p = .076, Greenhouse-Geisser correction). A two-tailed t-test conducted to compare unimodal auditory localization precision between groups (Fig. 3B) revealed a significant difference in localization precision between the control and amblyopia groups (t28 = 2.09, P = 0.046, d = .765).

Spearman rank correlations were conducted to assess the relation between localization precision and clinical measures of amblyopia. There were no significant relations between mean localization precision for visual blobs (collapsed across blob sizes) and amblyopic eye visual acuity (Rs = .14, P = 0.62) or stereo acuity (Rs = .45, P = 0.11). Similarly, there were no significant relations between localization precision for auditory clicks and amblyopic eye visual acuity (Rs = .099, P = 0.736) or stereo acuity (Rs = .054, P = 0.91).

#### Spatially Congruent Bimodal Stimuli

Localization precision for spatially congruent bimodal stimuli decreased in both the control group and amblyopia group as the blob size increased from 20°/SD to 28°/SD (Fig. 3C). A two (group) × four (blob size) mixed-design ANOVA for bimodal localization precision yielded a significant main effect of group (F1,28 = 6.724, P = 0.015, η̂2p = .194), a significant main effect of blob size (F1,98.55.39 = 34.42, P < 0.001, η̂2p = .551, Greenhouse-Geisser correction), but no significant interaction of group and blob size (F1,98.55.39 = 0.46, P = 0.650, η̂2p = .016, Greenhouse-Geisser correction).

Spearman rank correlations revealed no significant relations between mean localization precision for spatially congruent bimodal stimuli (collapsed across blob sizes) and amblyopic eye visual acuity (Rs = .42, P = 0.14) or stereo acuity (Rs = .005, P = 0.99).

### Agreement With MLE Model

According to the MLE model, audiovisual integration results in enhanced localization precision for bimodal stimuli by optimal combination of the component unimodal spatial signals. In complete integration failure, however, the best localization...
FIGURE 2. Unimodal and bimodal localization task performance and fitted cumulative normal functions for a representative participant in the control group (A, C, E, G) and the amblyopia group (B, D, F, H). Open symbols represent the proportion of trials in which the probe stimulus was perceived leftward of the test stimulus. Visual stimuli were Gaussian blobs of specified size (1 SD = 20°, red; 1 SD = 24°, orange; 1 SD = 28°, green; 1 SD = 32°, blue), and auditory stimuli were white noise clicks (black). (A, B) Unimodal conditions with visual or auditory test stimulus centered at 0°. (C, D) Bimodal conditions with unimodal components of test stimulus that in spatial agreement (i.e., blob and click both centered at 0°). (E-H) Bimodal conditions with unimodal components of test stimulus in spatial conflict to elicit the ventriloquist effect (i.e., blob and click symmetrically displaced 4° from center).
Optimal Audiovisual Spatial Integration in Amblyopia

Acuity and Performance with Intact Integration

The bimodal localization precision achievable is that of the more precise unimodal signal. This distinction provides a test for integration in amblyopia. Importantly, the MLE model also predicts that the bimodal enhancement in localization precision is greatest, and therefore most detectable, when the localization precisions of the unimodal components are equal (i.e., $\beta_V = \beta_A$) (see Equations 4 and 5 in Appendix A). The bimodal localization precision observed in this study was therefore compared with that expected with intact integration (i.e., MLE-predicted value computed from unimodal component precisions) and with integration failure (i.e., the most precise unimodal component) specifically for the condition in which the unimodal components were most similar for each participant (Fig. 4). For the control group, a one-way repeated measures ANOVA showed a significant difference between the observed precision, MLE-predicted precision, and best unimodal component precision ($F_{1,04,15.62} = 7.13, P = 0.016, \eta^2_p = 0.322, $ Greenhouse-Geisser correction). As expected, Bonferroni post hoc analysis revealed that the observed bimodal precision was significantly better than the best unimodal component precision ($P = 0.017, \eta^2_p = 0.325$), but not significantly different from the MLE-predicted precision ($P = 0.974, \eta^2_p < 0.001$). For the amblyopia group, the same one-way repeated measures ANOVA showed a significant difference between the observed precision, MLE-predicted precision, and best unimodal component precision ($F_{1,18,15.39} = 8.83, P = 0.007, \eta^2_p = 0.404, $ Greenhouse-Geisser correction). Bonferroni post hoc analysis revealed that the observed bimodal precision was significantly better than the best unimodal component precision ($P = 0.011, \eta^2_p = 0.400$), but not significantly different from the MLE-predicted precision ($P = 0.727, \eta^2_p = 0.010$). These findings indicate that bimodal localization precision is consistent with the MLE model of optimal integration both the control group and the amblyopia group.

Localization Bias

Unimodal Stimuli and Spatially Congruent Bimodal Stimuli. Localization bias was defined as the PSE of the fitted psychometric function. For unimodal stimuli, a two (group) $\times$ four (blob size) mixed-design ANOVA for PSE yielded no significant main effect of group ($F_{1,28} = 1.666, P = 0.494, \eta^2_p = 0.017$), no significant main effect of blob size ($F_{1,89,52.81} = 1.01, P = 0.367, \eta^2_p = 0.035, $ Greenhouse-Geisser correction), and no significant interaction of group and blob size ($F_{1,89,52.81} = 1.36, P = 0.295, \eta^2_p = 0.010, $ Greenhouse-Geisser correction). The mean PSE across blob sizes did not differ significantly from the expected value of 0° for the control group ($M = 0.17^\circ, 95\%$ confidence interval $[CI] = -0.28^\circ, 0.62^\circ$) or the amblyopia group ($M = -0.07^\circ, 95\% CI = 0.66^\circ, 0.52^\circ$).

For spatially congruent bimodal stimuli, a two (group) $\times$ four (blob size) mixed-design ANOVA for PSE yielded a significant main effect of group ($F_{1,28} = 0.24, P = 0.678, \eta^2_p = 0.006$), no significant main effect of blob size ($F_{2,24,62.58} = 2.22, P = 0.111, \eta^2_p = 0.074$).

Error bars: ± 1 SEM. For (A) the control group and (B) the amblyopia group, the observed bimodal precision was significantly better than that of the best unimodal component ($P < 0.05$), but not significantly different from that predicted by the MLE model.

![Figure 3](image-url) Mean localization precision for (A) visual-only, (B) auditory-only, and (C) spatially congruent bimodal audiovisual stimuli. Control group data are shown in blue and amblyopia group data are shown in red. Error bars: ± 1 SEM. Localization precision in the amblyopia group was significantly poorer than in the control group for unimodal visual, unimodal auditory, and bimodal audiovisual stimuli ($P < 0.05$).

![Figure 4](image-url) Comparison of observed and predicted bimodal localization precisions. MLE-predicted bimodal precisions were computed solely from unimodal precision measurements, and represent expected performance with intact integration. Best unimodal component precisions represent expected performance in integration failure. Error bars: ± 1 SEM. For (A) the control group and (B) the amblyopia group, the observed bimodal precision was significantly better than that of the best unimodal component ($P < 0.05$), but not significantly different from that predicted by the MLE model.
ventriloquism effect, with PSE biased toward the position of the visual blob) to audition-dominant (i.e., a reverse classic ventriloquism effect, with PSE biased toward the position of the auditory click) as the blob size increased from 20°/SD to 32°/SD (Fig. 5). A two (group) × four (blob size) mixed-design ANOVA for PSE yielded no significant main effect of group ($F_{2,24,62.58} = 3.32, P = 0.058, \eta^2_P = 0.106$, Greenhouse-Geisser correction). Bonferroni post hoc tests failed to detect the source of this interaction (see Supplementary Fig. S2). The mean PSE across blob sizes did not differ significantly from the expected value of 0° for the control group ($M = 0.19°$, 95% CI = $-0.09°$, 0.47°) or the amblyopia group ($M = 0.10°$, 95% CI = $-0.27°$, 0.47°).

**Bimodal Stimuli With Spatial Conflict.** Localization biases for bimodal stimuli with spatial conflict are illustrated in Figure 5 for the control group and the amblyopia group (individual data are shown in Figs. 2E–H). In these trials, the visual and auditory unimodal components were symmetrically displaced about 0°. Both groups showed a monotonic progression in localization bias from vision-dominant (i.e., a classic ventriloquism effect, with PSE biased toward the position of the visual blob) to audition-dominant (i.e., a reverse ventriloquism effect, with PSE biased toward the position of the auditory click) as the blob size increased from 20°/SD to 32°/SD (Fig. 5). A two (group) × four (blob size) mixed-design ANOVA for PSE yielded no significant main effect of group ($F_{1,28} = 1.04, P = 0.317, \eta^2_P = 0.036$), a significant main effect of blob size ($F_{3,84} = 106.30, P < 0.001, \eta^2_P = 0.792$), and no significant interaction of group and blob size ($F_{3,84} = 0.42, P = 0.738, \eta^2_P = 0.015$). This indicates that bimodal localization bias in the ventriloquism effect was not dissimilar between the two groups.

**Agreement With MLE Model**

According to the MLE model, the perceived location of a bimodal event in the ventriloquism effect is the weighted average of the unimodal visual and auditory position signals (see Equation 1 in Appendix A). The relative weights of vision and audition, in turn, are determined by their respective unimodal localization precisions (see Equations 2 and 3 in Appendix A). The observed and MLE-predicted bimodal localization biases may therefore be compared as an additional test of the MLE model. For the control group, a two (observed versus MLE-predicted) × four (blob size) repeated measures ANOVA for PSE yielded no significant difference between the observed and MLE-predicted biases ($F_{1,15} = 1.38, P = 0.259, \eta^2_P = 0.084$), a significant main effect of blob size ($F_{3,45} = 118.70, P < 0.001, \eta^2_P = 0.888$), and no significant interaction between the two factors ($F_{3,45} = 1.59, P = 0.206, \eta^2_P = 0.096$). Similarly, for the amblyopia group, a two (observed versus MLE-predicted) × four (blob size) repeated measures ANOVA for PSE yielded no significant difference between the observed and MLE-predicted biases ($F_{1,15} = 0.05, P = 0.826, \eta^2_P = 0.004$), a significant main effect of blob size ($F_{3,45} = 120.56, P < 0.001, \eta^2_P = 0.903$), and no significant interaction between the two factors ($F_{3,45} = 0.96, P = 0.410, \eta^2_P = 0.070$). These findings indicate that bimodal localization bias in the ventriloquism is consistent with the MLE model in both control group (Fig. 5A) and the amblyopia group (Fig. 5B).

**DISCUSSION**

We report that under binocular viewing conditions typical of everyday experience, amblyopia is associated with a pervasive impairment in spatial localization precision that involves visual, auditory, and audiovisual (i.e., multisensory) perception. Using the MLE model of the ventriloquism effect, however, we show that audiovisual spatial integration is intact in amblyopia, and that the abnormalities in audiovisual spatial perception are consistent with optimal sensory re-weighting in the face of unisensory deficits in both spatial vision and spatial hearing.

The unimodal visual localization task measured relative localization precision under binocular viewing conditions for diffuse visual blobs of various sizes. The large Gaussian blobs provided no high-contrast edges to aid in the location discrimination task. The negative relation between visual blob size and localization precision observed in both groups signifies that increasingly blurred visual targets elicit greater uncertainty in spatial perception. Compared with the control...
group, the amblyopia group showed a general reduction in visual localization precision across blob sizes despite normal visual acuity in the fellow eye, and presumably normal visual acuity under binocular viewing conditions. Several possibilities may account for this finding. Contrary to clinical dogma, vision in the fellow eye is not entirely normal. Careful psychophysical studies have shown that the fellow eye has reduced optotype and Vernier acuity, and perhaps more importantly, greater spatial uncertainty and distortion affecting both foveal and extra-foveal vision. Such deficits in spatial vision, affecting both the amblyopic and fellow eyes, may therefore contribute to the binocular visual localization deficit observed in the present study. Another possible contributing factor may be the temporal interval between the two stimuli whose positions were judged. Previous studies of spatial vision in the fellow eye (mentioned above) used static visual targets whose spatial elements were present simultaneously. In our study, however, the spatial elements (i.e., blobs) were separated by a temporal interval of 500 ms. Time-dependent factors such as reduced visual persistence or fixation instability in amblyopia may have therefore contributed to the observed deficit in localizing temporally separated visual targets. Specifically, reduced visual persistence may speed the rate of decay of the spatial signal, and fixation instability may jitter the mapping of physical space to retinal coordinates.

This study also revealed a novel amblyopic deficit in auditory spatial localization precision. Two features of the experimental task are particularly notable: (1) trials were conducted in darkness with no visual cues, and (2) localization did not involve pointing of any kind, but was a ‘left’ or ‘right’ determination entered as a button press on a gamepad. These features mean that the spatial uncertainty in amblyopic vision and visuomotor control cannot directly account for the observed unisensory auditory effect. Rather, they indicate that the sensory impairment in amblyopia extends beyond vision to the realm of spatial hearing. What might be the basis for the surprising association between amblyopia and impaired sound localization? Gori et al. have proposed that cross-sensory interactions during development serve a calibrating function such that the more robust and accurate sense informs the other. For example, in healthy children younger than 8 years of age, vision informs touch in spatial orientation discrimination, but touch informs vision in size discrimination. In early bilateral visual impairment, however, cross-sensory calibration is affected in a predictable way: haptic orientation discrimination (for which vision typically dominates) is impaired, but haptic size discrimination (for which touch typically dominates) is preserved. Impaired visual input in amblyopia may similarly affect the cross-sensory calibration of auditory spatial abilities in early life. What is striking about our results is that the impairment in cross-sensory calibration of auditory localization occurred despite access to normal visual acuity from the fellow eye, and presumably normal visual acuity under binocular viewing conditions of everyday experience.

We have conducted further experiments to specifically investigate the effects of abnormal vision on the calibration of auditory spatial localization during development, which will be the subject of a separate report.

The bimodal localization task measured relative localization precision under binocular viewing conditions for diffuse visual blobs of varying sizes paired with a simultaneous auditory click. As with unimodal stimuli, the amblyopia group showed a general impairment in bimodal localization precision across blob sizes. However, comparison with predictions of the MLE model showed that multisensory integration was intact in both the control and amblyopia groups. Indeed, the localization precision for spatially congruent bimodal stimuli (Fig. 4) and the localization bias for spatially incongruent bimodal stimuli in the ventriloquism effect (Fig. 5) represented optimal combination of the unimodal spatial features according to the MLE model. Overall, these findings provide independent validation for the MLE model of ventriloquism in a larger sample of typically sighted individuals, and suggest that at least some multisensory processing abnormalities reported in amblyopia do not reflect disordered multisensory integration, but rather unisensory deficits that feed into and propagate through an otherwise normal integrative network.

A common theme in studies of multisensory integration in children is that it develops relatively late compared with unisensory abilities. By some estimates, optimal multisensory integration does not arise until age 8 to 10 years, which is beyond the sensitive period for amblyopic visual loss, and after the age at which the greatest recovery may be achieved through occlusion or penalization therapy. This may explain why optimal integration in the ventriloquism effect is spared in amblyopia. If optimal integration typically arises after the amblyopic unisensory deficits are crystallized, then it will naturally be tuned to the stable audiovisual spatial capabilities of late childhood and beyond. This principle also implies that other multisensory perceptual anomalies in amblyopia (e.g., reduced susceptibility to the McGurk effect and poorer audiovisual asynchrony detection) may result from informational deficits at the unisensory level (e.g., increased spatial or temporal uncertainty) rather than deficits in multisensory integration.

**Acknowledgments**

Supported by Grant MOP 106663 from the Canadian Institutes of Health Research (CIHR; Ottawa, Ontario, Canada), Leaders Opportunity Fund from the Canada Foundation for Innovation (CFI; Ottawa, Ontario, Canada), the John and Melinda Thompson Endowment Fund in Vision Neurosciences at The Hospital for Sick Children (Toronto, Ontario, Canada), the Department of Ophthalmology and Vision Sciences at The Hospital for Sick Children (Toronto, Ontario, Canada), and the Vision Science Research Program at the University of Toronto (Toronto, Ontario, Canada).

Disclosure: M.D. Richards, None; H.C. Goltz, None; A.M.F. Wong, None

**APPENDIX A**

The MLE model has been put forward by several groups as a model for multisensory integration of spatial information involving vision. For the ventriloquism effect, the MLE model predicts that the perceived location of a bimodal event will be the weighted average of the locations of the unimodal events, such that:

\[
\hat{\Delta}_{VA} = w_{V} \hat{\Delta}_{V} + w_{A} \Delta_{A}
\]

where \(\hat{\Delta}_{V}\) and \(\hat{\Delta}_{A}\) are the unisensory localization estimates for vision and audition, \(w_{V}\) and \(w_{A}\) are the perceptual weights for vision and audition, and \(\Delta_{VA}\) is the resultant bimodal localization estimate. The perceptual weights, \(w_{V}\) and \(w_{A}\), sum to 1, and are inversely proportional the variances of the unisensory localization estimates, \(\sigma^{2}_{A}\) and \(\sigma^{2}_{V}\), such that:

\[
w_{V} = \frac{\sigma^{2}_{A}}{\sigma^{2}_{V} + \sigma^{2}_{A}}
\]

and

\[
w_{A} = \frac{\sigma^{2}_{V}}{\sigma^{2}_{A} + \sigma^{2}_{V}}
\]
Experimentally, localization variance can be estimated from the psychometric curve fit to the unimodal localization data. The combination of unisensory localization estimates in the MLE model is mathematically optimal in that it results in a bimodal localization estimate with the lowest possible variance (i.e., highest possible precision):

\[
\sigma_{VA}^2 = \frac{\sigma_v^2 \sigma_a^2}{\sigma_v^2 + \sigma_a^2} \leq \min(\sigma_v^2, \sigma_a^2)
\]

(4)

If the psychometric response is represented by a cumulative normal function, the variance of the function is inversely related to maximum slope, \( \beta \), at the inflection point of the curve:

\[
\beta = \left( \frac{1}{\sqrt{2\pi}} \right) \left( \frac{1}{\sqrt{\sigma^2}} \right)
\]

(5)

Therefore, following from Equations 4 and 5, the MLE model predicts that spatial localization precision (represented by \( \beta \)) for the bimodal event is greater than or equal to that of its unimodal components, and that the bimodal advantage in spatial localization precision is greatest when the precisions of the unisensory components are equal.

References


