Vision Research 157 (2019) 274–281

Contents lists available at ScienceDirect

Vision Research

journal homepage: www.elsevier.com/locate/visres

Short and long-term visual deprivation leads to adapted use of audiovisual information for face-voice recognition

Stefania S. Moro[a,b], Adria E.N. Hoover[a], Jennifer K.E. Steeves[a,b,*]

[a] Department of Psychology and Centre for Vision Research, York University, Toronto, Canada
[b] The Hospital for Sick Children, Toronto, Canada

ARTICLE INFO

Number of Reviews = 3

Keywords:
Monocular enucleation
Visual deprivation
Face
Voice
Object recognition
Audiovisual

ABSTRACT

Person identification is essential for everyday social interactions. We quickly identify people from cues such as a person's face or the sound of their voice. A change in sensory input, such as losing one's vision, can alter how one uses sensory information. We asked how people with only one eye, who have had reduced visual input during postnatal maturation of the visual system, use faces and voices for person identity recognition. We used an old/new paradigm to investigate unimodal (visual or auditory) and bimodal (audiovisual) identity recognition of people (face, voice and face-voice) and a control category, objects (car, horn and car-horn). Participants learned the identity of 10 pairs of faces and voices (Experiment 1) and 10 cars and horns (Experiment 2) and were asked to identify the learned face/voice or car/horn among 20 distractors. People with one eye were more sensitive to voice identification compared to controls viewing binocularly or with an eye-patch. However, both people with one eye and eye-patched viewing controls use combined audiovisual information for person identification more equally than binocular viewing controls, who favour vision. People with one eye were no different from controls at object identification. The observed visual dominance for binocular controls is larger for person compared to object identification, indicating that faces (vision) play a larger role in person identification and that person identity processing is unique from that for objects. People with long-term visual deprivation from the loss of one eye may have adaptive strategies, such as placing less reliance on vision to achieve intact performance, particularly for face processing.

1. Introduction

We typically experience a combination of different sensory stimuli at the same time each day. Social interaction is often an important part of one's day ranging from chatting with friends to identifying whether an approaching individual is a friend or foe. Multisensory cues, such as seeing a person's face and hearing their voice provide important information that contribute to distinguishing an individual person's identity. If all of our sensory systems are intact, we use them to our full advantage.

Humans with complete visual deprivation have shown evidence for changes in other sensory systems that indicate enhanced abilities with their remaining senses following a complete loss of a sensory system. For instance, congenitally blind individuals have shorter response times for auditory discrimination tasks (Röder et al., 1999), faster processing of language (Röder, Stock, Bien, Neville, Rösler, & 2002), enhanced sound localization (Lessard, Paré, Lepore, & Lassonde, 1998) and enhanced tactile perception (Sathian, 2000) compared to sighted individuals. This suggests underlying physiological changes within the systems responsible for these senses to support these behavioural enhancements. If visual input were reduced by half, as in people with one functioning eye, it seems reasonable to expect that the other intact sensory systems should function to the best of their ability in order to adapt and compensate for the partial loss of vision.

Monocular blindness, resulting from the surgical removal of one eye (enucleation) represents a unique human model for examining the consequences of the loss of binocularity. It is unlike other more common forms of monocular deprivation such as amblyopia or strabismus since the removal of the end organ eliminates all forms of visual input to the brain from that eye (Kelly, Moro, & Steeves, 2012). People with one eye have enhanced sound processing ability compared to controls (Hoover, Harris, & Steeves, 2012). When presented with auditory stimuli along the horizontal azimuth, people with one eye have consistently more accurate spatial localization within a field of +/- 78 degrees compared to control participants who were monocular viewing, binocular viewing or with their eyes closed (Hoover et al., 2012).

Typically, when we are simultaneously presented with auditory and visual stimuli, the visual information is processed preferentially over

* Corresponding author at: 1032 Sherman Health Science Research Centre, York University, 4700 Keele St., Toronto, ON, Canada.
E-mail address: steeves@yorku.ca (J.K.E. Steeves).

https://doi.org/10.1016/j.visres.2018.01.009
Received 27 July 2017; Received in revised form 11 January 2018; Accepted 14 January 2018
Available online 24 March 2018

auditory information. This is also known as the Colavita visual dominance effect (Colavita, 1974). People with one eye do not demonstrate the Colavita effect when asked to categorize rapidly presented audiovisual targets consisting of line drawings of common objects paired with common sounds (Moro & Steeves, 2012). Instead, people with one eye process auditory and visual components of audiovisual targets equally. Together with the enhanced auditory spatial localization (Hoover et al., 2012), these results suggest that auditory information may be processed more reliably for people with one eye and could be a form of sensory compensation for the loss of binocularity.

A common illusion occurs when the auditory and visual components of an audiovisual stimulus are spatially displaced relative to one another, resulting in the perception of a single event typically displaced towards the visual component, known as the ventriloquism effect (Welch & Warren, 1980). People with one eye show similar audiovisual localization for spatially disparate audiovisual stimuli compared to controls viewing binocularly or with one eye-patched (Moro, Harris, & Steeves, 2014). However, unlike binocular and eye-patched controls, people with one eye take longer to localize unimodal visual stimuli compared to unimodal auditory stimuli (Moro et al., 2014).

For the most part, spatial visual ability is intact or somewhat better than controls in people with the loss of one eye early in life despite a 50% reduction of visual input to the visual brain. For example, the ability to discriminate low-contrast global shape, a more complex visual process (hyperacuity), is enhanced in people with one eye compared to eye-patched controls and controls viewing dichoptically (Steeves, Wilkinson, González, Wilson, & Steinbach, 2004). Face processing, however, has emerged as an exception. People with one eye do not show the composite face effect and they take longer to process the shape and spacing between internal facial features (Kelly, Gallie, & Steeves, 2012). This mild impairment is face-specific and was not replicated when performing the same tasks on stimuli from other visual image categories, namely houses (Kelly et al., 2012).

Face perception is a unique and multi-faceted aspect of vision (Leopold & Rhodes, 2010). The ability to identify a person is facilitated when face information is integrated with voice information through crosstalk between the unimodal visual and unimodal auditory percepts (Campanella & Belin, 2008). Previous exposure to combined face-voice information during person identity encoding facilitates identification of that individual when only unimodal cues (face or voice) are available (Ellis, Jones, & Mosdell, 1997; Schweinberger, Herholz, & Sommer, 1997; Sheffert & Olson, 2004; von Kriegstein et al., 2008). Bimodal identity recognition shows interference effects (increased reaction time and decreased accuracy) due to the addition of auditory information to visual information when identifying bimodal (face-voice) stimuli (Joassin, Maurage, Bruyer, Crommelinck, & Campanella, 2004). This suggests that because unimodal face recognition is superior to voice recognition, the addition of voice information interferes with the efficient processing of the face (during bimodal face-voice pairings). When face stimuli are degraded and therefore less reliable relative to the voices, bimodal stimulus presentations led to an enhancement effect indicating that the more reliable sensory information (face or voice) has greater influence on person identity recognition (Joassin, Maurage, & Campanella, 2008).

Face-voice integration effects, similar to those found in healthy controls (Campanella & Belin, 2008), are seen in infants as young as 4 months of age (Bahrick, Netto, & Hernandez-Reif, 1998) and in non-human primates (Izumi & Kojima, 2004). This suggests that face-voice identification improves with development and experience as older infants (7 months of age) have better face-voice matching compared to younger infants (4 months of age) (Bahrick et al., 1998). People who have had one eye removed early in life experience monocular deprivation during postnatal visual system maturation. The long-term consequences of this abnormal visual experience during development may result in altered use of auditory and visual sensory information when tested later in life, as mature adults.

The current study investigates how people with one eye use auditory and visual information for person and object identity recognition compared to binocular and eye-patched viewing controls. Since people with one eye have half of the visual input to the brain, will this alter face-voice integration? Will audiovisual integration be affected more generally and also alter the identification of audiovisual objects (cars paired with horns)? Using the same paradigm as Hoover, Démonet, and Steeves (2010) we measure visual, auditory and audiovisual recognition of people and objects in people with one eye compared to eye-patched and binocular viewing controls. The addition of an eye-patched control group compares whether the effects observed in people with one eye are simply due to monocular viewing or whether they are the result of long-term visual deprivation from eye enucleation. In Experiment 1, we quantify person recognition for faces, voices and face-voice pair combinations. We predict that people with one eye will have poorer person identification compared to controls as a result of reduced sensitivity to faces since previously we have shown mild face perception impairments in this monocular group (Kelly et al., 2012). In Experiment 2, we quantify object recognition for cars, car horns and car-car horn pair combinations. We predict no difference in object identification between groups. We have previously shown that higher-level aspects of spatial form vision are intact (Kelly et al., 2012) or somewhat enhanced (Steeves et al., 2004) for this group compared to eye-patched viewing. We also compare across experiments to assess whether person recognition is unique compared to object recognition.

2. Experiment 1: Person identity recognition

2.1. Materials and methods

2.1.1. Participants

2.1.1.1. People with one eye (monocular enucleation, ME). Eleven adult participants who had undergone early monocular eye enucleation (ME) at The Hospital For Sick Children (Toronto) participated in this study (mean age = 34 years, SD = 12). All ME participants had been unilaterally eye enucleated (7 right eye removed) due to retinoblastoma, a rare childhood cancer of the retina. Age at enucleation ranged from 4 to 66 months of age (mean age at enucleation = 22 months, SD = 16).

2.1.1.2. Binocular viewing control participants (BV). Twenty-five binocularly intact controls with a mean age of 27 years (SD = 7) were tested while viewing stimuli binocularly.

2.1.1.3. Patched viewing control participants (MV). Twenty-five binocularly intact participants, with a mean age of 24 years (SD = 3), completed the experiments with one eye patched. Participants' non-preferred eye (determined using the Porta test) was patched with a semi-opaque eye covering and translucent tape (12 right-eye covered).

All participants (ME, BV, MV) reported normal hearing and normal or corrected-to-normal acuity as assessed by an EDTRS eye chart (Precision Vision™, La Salle, IL) and wore optical correction if needed. All participants gave informed consent prior to their inclusion in the study, which was approved by York University Office of Research Ethics.

2.1.2. Stimuli

Stimuli were previously used in Hoover et al. (2010) for assessing face and object identities in a patient with visual agnosia. See Hoover et al. (2010) for more detailed information on how stimuli were created. In short, visual stimuli consisted of 110 greyscale female face images that were cropped within an oval aperture and distinguishable features such as beauty marks removed. Auditory stimuli were 20 s in duration and were played through headphones. Each consisted of a short neutral passage spoken in English by one of 110 female voices.

Participants sat 45 cm from the display in a dimly lit room with the

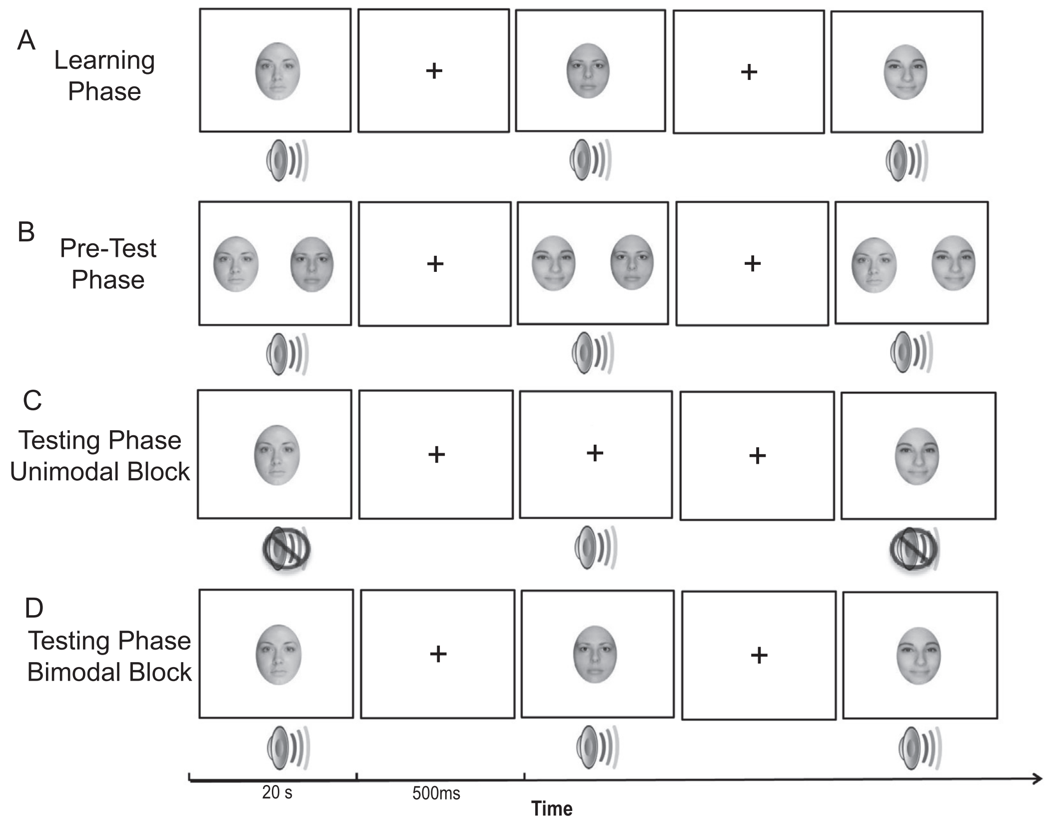

Fig. 1. Schematic diagram representing the task in Experiment 1. A. Learning phase. B. Pre-test phase. C. Testing Phase Unimodal Block. D. Testing Phase Bimodal Block.

head stabilized by a chin/forehead rest. All stimuli were presented with Superlab 4.0 (Cedrus, Inc.) and participants indicated their responses using designated keys on a keyboard.

2.1.3. Procedure

The current experimental procedure follows that used by Hoover et al. (2010). *Learning phase:* Participants learned the identities of 10 individuals who were previously unknown. Each "identity" consisted of a visual face image paired with a voice stimulus. Participants were told that they would be tested on their ability to recognize these ten identities and that they had to learn to associate each specific face with its corresponding voice. Participants viewed a static face image and listened to its paired voice for 20 s. A fixation-cross (500 ms in duration) appeared before the visual stimuli (Fig. 1A). Each of the ten identities were presented four times. In total, the learning phase lasted approximately 15 min.

Pre-test phase: Following the learning phase participants were given a short pre-test to ensure an adequate level of learning of the paired face/voice identities. During this phase participants heard one of the learned voices and saw two of the learned faces displayed side by side, one of which corresponded to the voice. Participants were instructed to press a button on the response pad corresponding to which of the two faces matched the voice that they heard. Participants were unable to advance to the next trial until they responded correctly. The same trial was repeated immediately after the incorrect trial until the correct response was given. Upon selection of a correct response the trial would end and proceed to the following one (Fig. 1B). Each voice was initially presented twice to yield a minimum of 20 trials. In total, the pre-test phase lasted approximately 7 min.

Testing phase: This phase followed the pre-test phase and consisted of two different blocks (unimodal visual or auditory) and bimodal (face-voice pairs). During the unimodal block participants viewed a visual face image alone or heard a voice stimulus alone and were asked to identify as quickly and accurately as possible whether the identity of the stimulus was learned or new. The ten learned face or voice identities were presented twice among 20 new (unlearned) distractor faces and 20 new voices. The unimodal block consisted of 40 face only trials and 40 voice only trials for a total of 80 trials presented in random order. In the bimodal block participants heard a voice and saw a face at the same time. Participants were asked to identify as quickly and accurately as possible whether the face-voice identity was learned or new using a designated button on a response pad. A correct learned identity consisted of a congruent presentation of a previously learned face and voice. 10 learned identities were presented twice each among 30 new distractor identities for a total of 50 trials presented in random order. The faces and voices were paired in four different combinations: (1) a learned face with a learned voice (congruent), (2) a learned face with a new voice (incongruent), (3) a new face with a learned voice (incongruent), (4) a new face with a new voice (congruent) (Fig. 1C). In total, the testing phase lasted a maximum of approximately 40 min.

2.2. Results

Sensitivity (*d′*) scores were calculated for all conditions (visual only, auditory only, overall bimodal, congruent bimodal and incongruent bimodal) for all participants. The overall bimodal sensitivity scores were calculated based on both congruently and incongruently paired stimuli. Some participants had perfect performance, which was addressed by adjusting the values according to the method recommended by Macmillan and Creelman (1991) for perfect scores.

To evaluate bimodal enhancement (or interference) we calculated bimodal effect scores by subtracting the unimodal sensitivity score, either visual or auditory from the congruent bimodal score. To evaluate the effect of a unimodal stimulus preference on bimodal performance congruency effects with the bimodal pairings were investigated. We calculated a congruency difference score by subtracting the incongruent bimodal sensitivity (false alarms from the incongruent pairings were used in the calculation of the *d′* score) score from the congruent bimodal sensitivity score. In general the congruency effects allow us to evaluate the presence of a preference for unimodal stimuli (vision or audition) in overall bimodal identity formation. For example, if binocular viewing controls make their recognition judgments based on their dominant modality (vision) then we would predict they would respond that a learned face image paired with a new voice is a “learned” identity.

2.2.1. Overall identity recognition performance

A 3 × 3 repeated measures analysis of variance (ANOVA) was conducted to compare sensitivity scores for overall identity recognition performance in each participant group (ME vs BV vs MV) and stimulus condition (auditory only, visual only, overall bimodal). There was a significant interaction, $F(4, 116) = 5.228$, $p = 0.001$, $\eta_p^2 = 0.153$ and main effect of condition, $F(2, 116) = 49.019$, $p < 0.001$, $\eta_p^2 = 0.458$. There was a trending main effect of participant group, $F(2, 58) = 3.055$, $p = 0.055$, $\eta_p^2 = 0.095$. Bonferroni corrected post hoc tests for the main effect of condition indicated a significant heightened sensitivity to voice only stimuli compared to the BV group for the MV group ($p = 0.003$) and ME group ($p = 0.008$). Bonferroni corrected post hoc tests were conducted for the interaction comparing the sensitivity scores in each modality for each participant group. The sensitivity scores for the BV group visual (face only) was greater compared to both auditory (voice only) ($p < 0.001$) and bimodal (face/voice) ($p < 0.001$). Additionally the bimodal (face/voice) was greater compared to the auditory (voice only) ($p < 0.001$). The sensitivity scores for the MV group visual (face only) was greater compared to the auditory (voice only) ($p < 0.001$) but not bimodal (face/voice) ($p = 0.150$). Additionally the bimodal (face/voice) was greater compared to the auditory (voice only) ($p = 0.004$). The sensitivity scores for the ME group visual (face only) was greater compared to the auditory (voice only) ($p = 0.029$) but not compared to bimodal (face/voice) ($p = 1.000$). Additionally there was no difference in bimodal (face/voice) compared to the auditory (voice only) ($p = 0.270$). Fig. 2 plots the sensitivity scores for each for the BV, MV and ME groups for unimodal auditory, unimodal visual and bimodal conditions.

2.2.2. Bimodal effect scores

Visual effect score: A one way ANOVA comparing sensitivity to unimodal visual stimuli compared to bimodal stimuli (congruent bimodal sensitivity – voice only sensitivity) in each participant group (ME vs BV vs MV) was not significant, $F(2, 60) = 2.114$, $p = 0.130$. The ME group, MV group and BV group did not differ in sensitivity when faces and voices were paired together compared to when voices were tested alone. One sample t-tests were conducted to determine whether the bimodal effect scores for each participant group indicated a significant enhancement effect (significantly different from zero). A significant enhancement effect was found for the BV ($t(24) = 3.523$, $p = 0.025$) and the MV group ($t(24) = 2.443$, $p = 0.022$). No significant enhancement was found for the ME group ($t(10) = 1.669$, $p = 0.126$). Fig. 3 plots the difference in sensitivity (d′) of the visual effect scores for each for the BV, MV and ME groups.

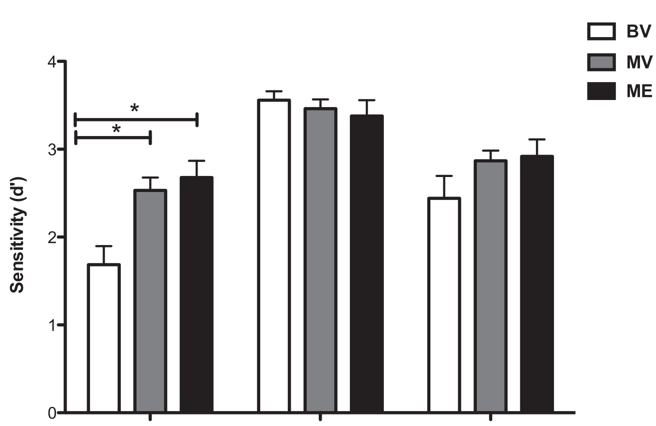

Fig. 2. The sensitivity scores (d′) for overall identity recognition for each of the BV (white), MV (grey) and ME groups (black) for the unimodal auditory, unimodal visual and bimodal conditions. Significant differences in sensitivity are indicated by * ($p < 0.05$). Error bars represent standard error of the mean.

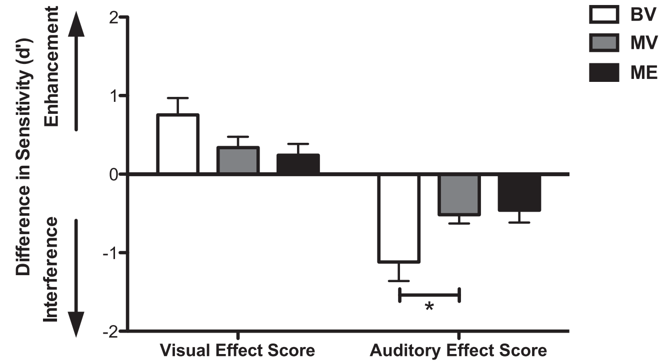

Fig. 3. Visual effect scores: congruent bimodal-unimodal auditory (first three columns) and the auditory effect scores: congruent bimodal-unimodal visual (last three columns) for each of the BV (white), MV (grey) and ME (black) groups. Significant differences are indicated by * ($p < 0.05$). Error bars represent standard error of the mean.

Auditory effect score: A one way ANOVA comparing the effect of sensitivity to unimodal auditory stimuli compared to bimodal stimuli (congruent bimodal sensitivity – face only sensitivity) in each participant group (ME vs BV vs MV) was significant, $F(2, 60) = 3.665$, $p = 0.032$. Bonferroni corrected post hoc tests indicate that the BV group demonstrated increased interference compared to the MV group ($p = 0.05$) when faces and voices were paired together compared to when faces were tested alone. One sample t-tests were conducted to determine whether the bimodal effect scores from each participant group indicated a significant interference effect (significantly different from zero). A significant interference effect was found for all participant groups, BV ($t(24) = -4.640$, $p < 0.001$), MV ($t(24) = -4.689$, $p < 0.001$) and ME groups ($t(10) = -2.894$, $p = 0.016$). Fig. 3 plots the difference in sensitivity (d′) of the auditory effect scores for each for the BV, MV and ME groups.

2.2.3. Congruency difference scores

A one way ANOVA comparing sensitivity to congruency (congruent bimodal sensitivity – incongruent bimodal sensitivity) in each participant group (ME vs BV vs MV) was significant, $F(2, 60) = 6.873$, $p = 0.002$. Bonferroni corrected post hoc tests indicated that the BV group had higher sensitivity compared to the MV group ($p = 0.014$) and ME group ($p = 0.006$) when learned faces and voices were paired together congruently compared to an incongruent pairing (a learned face or voice paired with a new voice or face). One sample t-tests were conducted to determine whether each participant group demonstrated a preference for congruently paired bimodal stimuli (significantly greater congruency difference compared to zero). A significant preference for congruently paired bimodal stimuli was found for the BV group, BV ($t(24) = 4.168$, $p < 0.001$). No significant differences were found for the MV ($t(24) = 0.794$, $p = 0.435$) and ME groups ($t(10) = -0.908$, $p = 0.385$). Fig. 4 plots the congruency difference scores for each of the BV, MV and ME groups.

2.3. Summary of person identity recognition results

In Experiment 1, people with one eye and eye-patched controls were more sensitive to unimodal voice stimuli compared to binocular

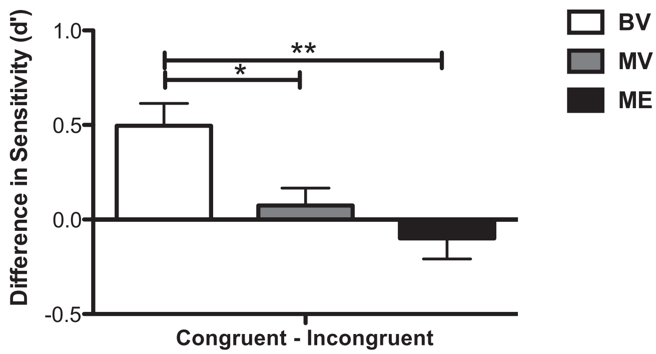

Fig. 4. Difference in sensitivity (d′) for the congruency difference scores plotted for each of the BV (white), MV (grey) and ME (black) groups. Significant differences are indicated by * ($p < 0.05$) and ** ($p < 0.001$). Error bars represent standard error of the mean.

viewing participants (Fig. 2, auditory sensitivity). This suggests an adaptation to the reduction of visual input from one eye in general. All three participant groups demonstrated a significant interference effect when their bimodal sensitivity score was compared to their unimodal visual sensitivity score (Fig. 3, auditory effect score). Binocular viewing controls showed significantly greater interference in their auditory effect score compared to eye-patched viewing controls and moreover showed greater (but not significant) interference compared to people with one eye (Fig. 3, auditory effect score). This indicates that the additional information from the voice hindered identification for the binocular viewing controls. People with one eye and eye-patched controls showed greater sensitivity to voices compared to the binocular controls, and the addition of the voice information to a visual face did not change sensitivity as much as it did for binocular viewing controls who had poor voice sensitivity.

People with one eye did not show a difference in performance compared to binocular and eye-patched viewing controls when their bimodal sensitivity score was compared to their unimodal auditory sensitivity score (Fig. 3, visual effect score). Furthermore, people with one eye did not show a significant enhancement in bimodal sensitivity, unlike both control groups. This indicates that, despite not demonstrating an enhancement effect, the additional information from their partially deprived sense did not hinder their performance as they did not differ from either control group.

Binocular viewing controls demonstrated a preference for congruently paired bimodal stimuli while both eye-patched controls and people with one eye were not affected to the same extent by the incongruently paired bimodal stimuli (Fig. 4). This suggests that binocular viewing controls likely made their recognition judgments based on their dominant sensory modality (vision) more often than eye-patched controls and people with one eye. This conclusion can be made by evaluating the data in conjunction with binocular viewing controls' increased enhancement of their visual effect score and increased interference in their auditory effect score.

3. Experiment 2: Object identity recognition

In Experiment 2, we sought to determine whether these audiovisual differences were specific to person identity or whether they occurred more generally for other audiovisual stimuli. We repeated these experiments with an object stimulus: cars and horns.

3.1. Materials and methods

3.1.1. Participants

The same participants completed Experiment 1 and Experiment 2.

3.1.2. Stimuli

Stimuli consisted of 110 greyscale images of cars taken from the

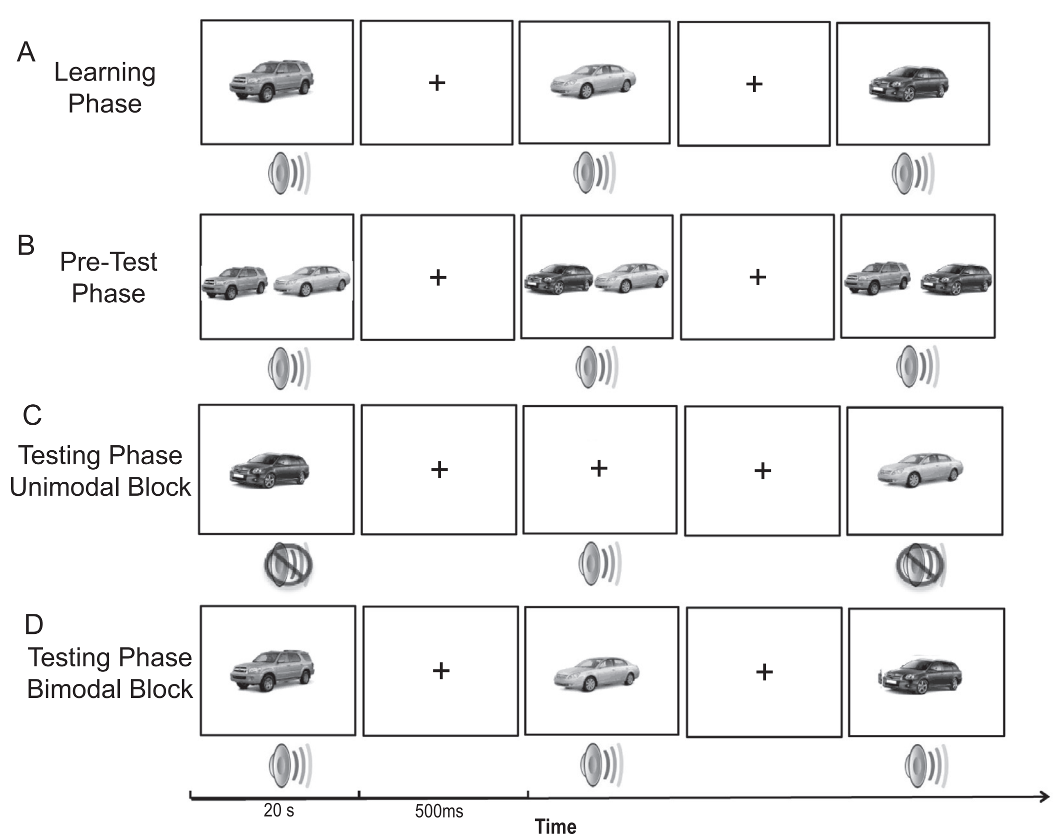

Fig. 5. Schematic diagram representing the task in Experiment 2. A. Learning phase. B. Pre-test phase. C. Testing Phase Unimodal Block. D. Testing Phase Bimodal Block.

internet and resized (see Hoover et al. 2010). Auditory horn stimuli were 20 s in duration and consisted of 110 different car horn sounds (see Hoover et al. 2010).

3.1.3. Procedure

As in Experiment 1, participants sat 45 cm from the display in a dimly lit room with the head stabilized by a chin/forehead rest. The procedure was identical to that for Experiment 1 throughout substituting car images for faces and car horns for voices (Fig. 5).

3.2. Results

The same analyses were used as in Experiment 1.

3.2.1. Overall identity recognition

A 3 × 3 repeated measures analysis of variance (ANOVA) comparing sensitivity scores for overall object identity recognition performance in each participant group (ME vs BV vs MV) and stimulus condition (auditory only, visual only, overall bimodal) revealed no significant interaction, $F(4, 116) = 1.501$, $p = 0.207$, $\eta_p^2 = 0.049$ and main effect of participant group, $F(2, 58) = 0.705$, $p = 0.498$, $\eta_p^2 = 0.024$. There was a main effect of condition, $F(2, 116) = 17.708$, $p < 0.001$, $\eta_p^2 = 0.234$. Bonferroni corrected post hoc tests indicated that the participant groups were less sensitive to auditory stimuli (horns) compared to visual stimuli (cars) ($p = 0.005$) and bimodal (cars/horns) ($p = 0.048$). The ME, MV and BV groups did not demonstrate a difference in sensitivity across groups within any of the conditions. Fig. 6 plots the sensitivity scores for each of the BV, MV and ME groups.

3.2.2. Bimodal effect scores

Visual effect score: A one way ANOVA comparing sensitivity to unimodal visual stimuli compared to bimodal stimuli (congruent bimodal sensitivity – horn only sensitivity) in each participant group (ME vs BV vs MV) was not significant, $F(2, 60) = 1.266$, $p = 0.290$. The ME group, MV group and BV group did not differ in sensitivity when cars and horns were paired together compared to when horns were tested alone. One sample t-tests were conducted to determine whether the bimodal effect scores from each participant group indicated a significant enhancement effect (significantly different from zero). A significant enhancement effect was found for the MV ($t(24) = 2.548$, $p = 0.018$) and the ME group ($t(10) = 2.890$, $p = 0.016$). No significant enhancement was found for the BV group ($t(24) = 1.482$, $p = 0.151$). Fig. 7 plots the difference in sensitivity (d′) of the visual effect scores for each of the BV, MV and ME groups.

Auditory effect score: A one way ANOVA comparing sensitivity to unimodal auditory compared to bimodal sensitivity (congruent bimodal sensitivity – car only sensitivity) in each participant group (ME vs BV vs MV) was also not significant, $F(2, 60) = 2.770$, $p = 0.071$. The ME, MV and BV groups did not differ in sensitivity when cars and horns were paired together compared to when horns were tested alone. One sample t-tests were conducted to determine whether the bimodal effect scores from each participant group indicated a significant interference effect (significantly different from zero). A significant interference effect was found for the BV ($t(24) = -3.681$, $p = 0.001$). No significant interference was found for the MV group ($t(24) = -0.579$, $p = 0.568$) and the ME group ($t(10) = -0.014$, $p = 0.989$). Fig. 7 plots the difference in sensitivity (d′) of the auditory effect scores for each of the BV, MV and ME groups.

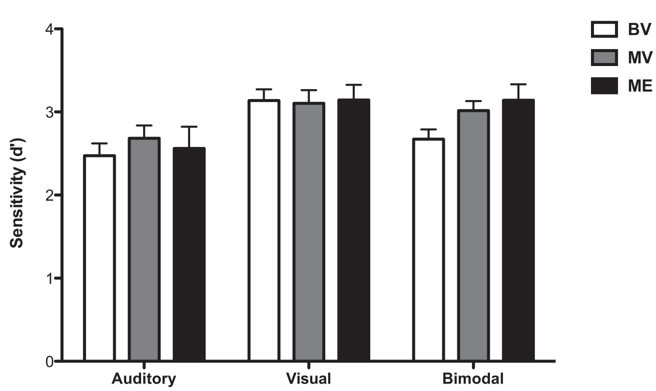

Fig. 6. Sensitivity scores (d′) for each of the BV (white), MV (grey) and ME groups (black) for the unimodal auditory, unimodal visual and bimodal conditions. Error bars represent standard error of the mean.

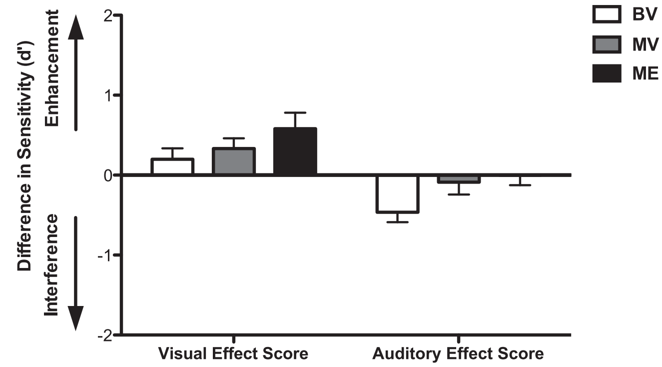

Fig. 7. Visual effect scores: congruent bimodal-unimodal auditory (first three columns) and the auditory effect scores: congruent bimodal-unimodal visual (last three columns) for each of the BV (white), MV (grey) and ME (black) groups. Error bars represent standard error of the mean.

3.2.3. Congruency differences

A one way ANOVA comparing sensitivity to congruency (congruent bimodal sensitivity – incongruent bimodal sensitivity) in each participant group (ME vs BV vs MV) was not significant, $F(2, 60) = 3.031$, $p = 0.056$. The BV group, MV group and ME group showed no difference in sensitivity when learned cars and horns were paired together congruently compared to an incongruent pairing made up of a learned car or horn and a new car or horn. One sample t-tests were conducted to determine whether each participant group demonstrated a preference for congruently paired bimodal stimuli (significantly greater congruency difference compared to zero). A significant preference for congruently paired bimodal stimuli was found for the BV group, BV ($t(24) = 3.676$, $p = 0.001$). No significant differences were found for the MV ($t(24) = 0.748$, $p = 0.462$) and ME groups ($t(10) = -0.278$, $p = 0.787$). Fig. 8 plots the congruency difference scores for each of the BV, MV and ME groups.

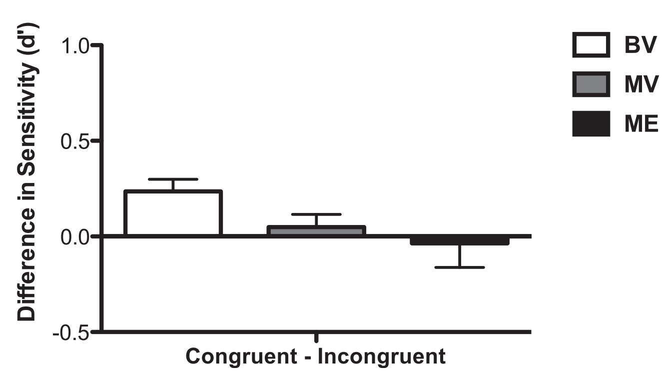

Fig. 8. Difference in sensitivity (d′) for the congruency difference scores plotted for each of the BV (white), MV (grey) and ME (black) groups. Error bars represent standard error of the mean.

3.3. Summary of object identity recognition results

In Experiment 2, there was no difference in performance between groups for unimodal visual (car), unimodal auditory (horn) and bimodal (car-horn) stimuli sensitivity (Fig. 6). This indicates that the auditory advantage demonstrated by people with one eye for auditory only stimuli is specific to person identification and does not generalize to object identification. The eye-patched control group and people with one eye demonstrated a significant enhancement effect when their bimodal sensitivity score was compared to their unimodal auditory sensitivity score (Fig. 7, visual effect score). The binocular viewing control group demonstrated a significant interference effect when their bimodal sensitivity score was compared to their unimodal visual sensitivity score (Fig. 7, auditory effect score). Overall, when comparing between participant groups however, people with one eye have no increased or decreased interference or enhancement for either of their bimodal sensitivity scores (Fig. 7). All three groups did not differ in identifying congruently paired compared to incongruently paired bimodal stimuli (Fig. 8).

3.4. Comparison across face-voice and car-horn tasks

Visual effect score – Face-voice vs. *Car-horn:* Paired samples t-tests were conducted to compare the visual effect scores obtained in each task (face-voice and car-horn) for each participant group (ME, BV and MV). A significant difference in visual effect scores was found across tasks for the BV group, $t(24) = 2.093$, $p = 0.047$. No significant differences in visual effect scores were found across tasks for both the MV group, $t(24) = 0.027$, $p = 0.979$ and ME group, $t(10) = -1.326$, $p = 0.214$. Fig. 9 plots the visual effect scores for each task for the BV, MV and ME groups.

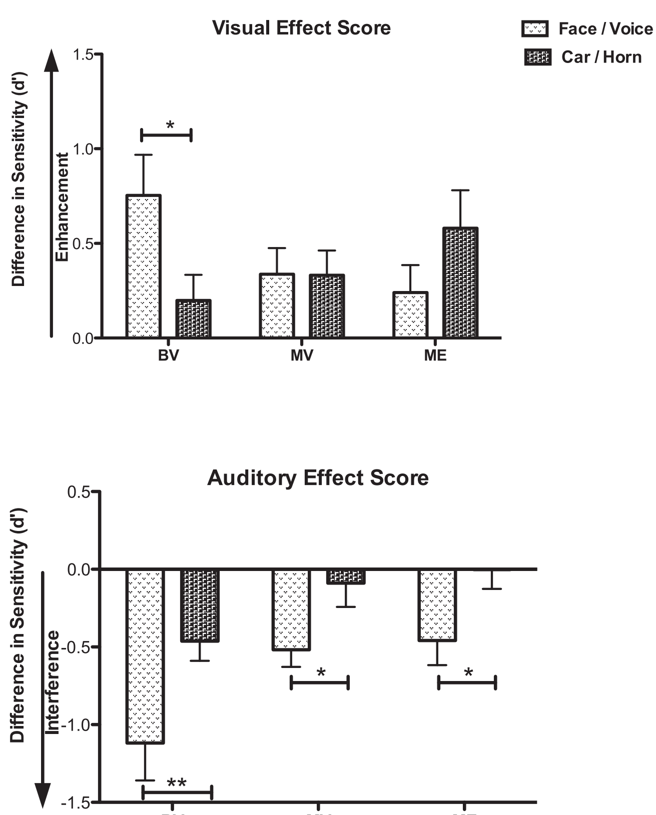

Fig. 9. The visual effect score and auditory effect score for each task for each of the BV, MV and ME groups comparing across the face-voice (white) and car-horn stimuli (black) tasks. Significant differences are indicated by * ($p < 0.05$). Error bars represent standard error of the mean.

Auditory effect score – Face-voice vs. *Car-horn:* Paired samples t-tests were conducted to compare the auditory effect scores obtained in each task (face-voice and car-horn) for each participant group (ME, BV and MV). A significant difference between auditory effect scores was found across tasks for all of the participant groups: BV group, $t(24) = -2.440$, $p = 0.022$, MV group, $t(24) = -2.193$, $p = 0.038$ and ME group, $t(10) = -2.249$, $p = 0.048$. Fig. 9 plots the auditory effect scores for each task for the BV, MV and ME groups.

4. General discussion

The current study investigated how people with one eye use auditory and visual information for person and object identity recognition compared to binocular viewing and eye-patched controls. When comparing across tasks, only the binocular viewing controls' sensitivity to person recognition was enhanced by adding the visual stimulus to the auditory stimulus (visual effect score) (Fig. 9, BV group). This effect was limited to the person identity but not the object identity task. These results indicate a form of visual dominance for faces consistent with other visual dominance effects such as the Colavita effect (Moro & Steeves, 2012; 2013). Binocular viewing controls use visual information from faces to enhance person identification more than the visual information from cars to enhance object identification. Eye-patched viewing controls and people with one eye did not differ across tasks on their visual effect score suggesting a lack of visual dominance, consistent with previous research (Moro & Steeves, 2012). Additional information provided by the visual stimulus did not enhance bimodal sensitivity for either the eye-patched control group or people with one eye, and had little effect on performance in the audiovisual condition for either person and object identity. This is likely due to a ceiling effect given the high sensitivity for the unimodal condition. All groups showed greater interference on their auditory effect score, where additional information provided by the auditory stimulus more greatly interfered with bimodal sensitivity for the face-voice task (Experiment 1). However, only the eye-patched viewing controls and people with one eye showed interference as per their auditory effect score for the car-horn task (Experiment 2). This result indicates that the additional auditory information from the voice hindered the sensitivity for person identification greater than the additional auditory information from the horns for object identification.

The binocular viewing controls have consistent performance with those from other face-voice person recognition studies (Hoover et al., 2010; Joassin et al., 2008). Unimodal face recognition is superior to voice recognition and the addition of voice information interferes with the efficient processing of the face during bimodal face-voice pairings in binocular viewing controls (Joassin et al., 2008). In the current study, people with one eye demonstrate increased sensitivity to faces compared to voices, similar to binocular viewing controls however, they do not show increased sensitivity to faces compared to bimodal face-voice stimuli. People with one eye have mild impairments (are slower) in processing the shape and spacing between internal facial features indicating an overall subtle deficit in processing whole faces (Kelly et al., 2012). This deficit, although minor, might play a more substantial role in deciphering person identity and result in people with one eye taking greater advantage of available voice information for bimodal person identity recognition.

In the current study, the eye-patched viewing control group performs more similarly to the people with one eye rather than the binocular viewing control group. Our lab has previously found effects of monocular viewing from eye patching on low level, low contrast visual tasks (Reed, Steeves, Kraft, Gallie, & Steinbach, 1996; Steeves et al., 2004) and this is the first time that an effect of eye patching has been observed on higher level visual processing. The similar performance can be attributed to the overall reduction of visual input to the brain whether short-term from the eye-patch or long-term from early eye enucleation in childhood. Despite the reduction of visual input, bimodal person identification is nonetheless intact in both groups. These results support a top-down model of face and voice encoding where identification is modulated by supramodal brain regions rather than a bottom up model where extensive unisensory processing takes place to support supramodal brain regions (Maguinness & von Kriegstein, 2017; Latinus, VanRullen, & Taylor, 2010). Person identity performance may be affected by a reduction of visual input, however, it is important to note that the type of visual deprivation varies substantially between the two visually deprived groups in this study. On one hand, people with one eye likely undergo long-term developmental changes following early eye-enucleation that results in permanent cortical reorganization (Kelly, DeSimone, Gallie, & Steeves, 2015; Moro, Kelly, McKetton, Gallie, & Steeves, 2015) that contributes to altered underlying face-voice processing mechanisms. On the other hand, eye-patched viewing controls only experience short-term visual deprivation at the time of testing which appears to transiently alter the use or weighting of auditory information during face/voice recognition.

People with one eye have better sensitivity to voices compared to binocular controls. In accordance with the modality appropriateness hypothesis, the best sensory modality at processing a specific sensory dimension or stimulus attribute, should dominate (Welch & Warren, 1980). It is possible that the increased sensitivity in hearing voices allows for more use of voices for identity recognition and serves as an adaptive compensatory mechanism for the loss of one half of the visual input to the brain. The result is a more "balanced" contribution of each sense for bimodal (audiovisual) person identification and not a dominance of one sense (typically vision) over another.

The increased reliance on visual face information in all three participant groups (although greater for binocular viewing controls and eye-patched viewing controls) for person identification compared to object identification suggests that person identification is a unique process that is in some ways distinct from object identification. It has been shown that learned face-voice stimuli have underlying neural circuits that work in conjunction with one another (FFA, fusiform face area and TVA, temporal voice area) when comparing between sensory redundant stimuli (faces and voices) or arbitrary multimodal combinations (cell phones and ring tones) (von Kriegstein & Giraud, 2006). Information generated from person identification has stronger, more meaningful and ecological associations than those for object identification (Hoover et al., 2010). Visual face information, therefore, facilitates person identity recognition when presented in conjunction with a person's voice. People with one eye, despite demonstrating a difference between tasks, may still use audiovisual information equally for object identification and person identification, since they are better at voice recognition compared to binocular controls. In addition to demonstrating mild impairments in holistic face processing (Kelly et al., 2012) people with one eye lack typical visual dominance for non-face stimuli (Moro & Steeves, 2012) perhaps indicating more generalized decrease in visual dominance across higher level visual stimuli (such as both faces and objects). It has recently been shown that people with one eye have increases in surface area and gyrification in auditory, and multisensory cortices in addition to the visual cortex (Kelly et al., 2015), as well as, a volume asymmetry in the MGB that may represent increased interactions between the left MGB and primary auditory cortex (Moro et al., 2015). These subcortical and cortical adaptations could support altered audiovisual contributions to person and object identification mechanisms.

In conclusion, short-term visual deprivation (from eye-patching) and long term visual deprivation (from removal of an eye early in life) decrease the reliability of face stimuli and may yield better use of voice information during face-voice person identification. People with one eye demonstrate evidence for generalized altered audiovisual neural contributions to person identification that may serve as an adaptive compensatory mechanism for the loss of half of the visual input to the brain.

Acknowledgments

We sincerely thank all of our participants for taking part in this study. We are also grateful to Adi Sheinberg and Ben Shachar for their assistance in testing participants. This research was supported by grants from the Natural Sciences and Engineering Research Council of Canada, as well as, the Canada Foundation for Innovation to JKES. SSM is supported by scholarships from the Natural Sciences and Engineering Research Council of Canada, Toronto Rehab's TD Graduate Scholarship for People with Disabilities and York University.

References

Bahrick, L. E., Netto, D., & Hernandez-Reif, M. (1998). Intermodal perception of adult and child faces and voices by infants. *Child Development, 69*, 1263–1275.

Campanella, S., & Belin, P. (2008). Integrating face and voice in person perception. *Trends in Cognitive Neuroscience, 11*, 535–543.

Colavita, F. B. (1974). Human sensory dominance. *Perception and Psychophysics, 16*, 409–412.

Ellis, H. D., Jones, D. M., & Mosdell, N. (1997). Intra- and inter modal repetition priming of familiar faces and voices. *British Journal of Psychology, 88*, 143–156.

Hoover, A. E. N., Démonet, J.-F., & Steeves, J. K. E. (2010). Superior voice recognition in a patient with acquired prosopagnosia and object agnosia. *Neuropsychologia, 48*, 3725–3732.

Hoover, A. E. N., Harris, L. R., & Steeves, J. K. E. (2012). Sensory compensation in sound localization in people with one eye. *Experimental Brain Research, 216*, 565–574.

Izumi, A., & Kojima, S. (2004). Matching vocalizations to vocalizing faces in a chimpanzee (Pan troglodytes). *Animal Cognition, 7*, 179–184.

Joassin, F., Maurage, P., Bruyer, R., Crommelinck, M., & Campanella, S. (2004). When audition alters vision: An event-related potential study of the cross-modal interactions between faces and voices. *Neuroscience Letters, 369*, 132–137.

Joassin, F., Maurage, P., & Campanella, S. (2008). Perceptual complexity of faces and voices modulates cross-modal behavioural facilitation effects. *Neuropsychological Trends, 3*, 29–44.

Kelly, K. R., Gallie, B. L., & Steeves, J. K. E. (2012). Impaired face processing in early monocular deprivation from enucleation. *Optometry and Vision Science, 89*, 137–147.

Kelly, K. R., Moro, S. S., & Steeves, J. K. E. (2012). Living with one eye: Plasticity in visual and auditory systems. In J. K. E. Steeves, & L. R. Harris (Eds.). *Plasticity in sensory systems* (pp. 94–108). Cambridge, UK: Cambridge University Press.

Kelly, K. R., DeSimone, K. D., Gallie, B. L., & Steeves, J. K. E. (2015). Increased cortical surface area and gyrification following long-term survival from early monocular enucleation. *NeuroImage: Clinical, 7*, 297–305.

Latinus, M., VanRullen, R., & Taylor, M. J. (2010). Top-down and bottom-up modulation in processing bimodal face/voice stimuli. *BioMed Central Neuroscience, 11*, 1–13.

Leopold, D. A., & Rhodes, G. (2010). A comparative view of face perception. *Journal of Comparative Psychology, 124*, 233–251.

Lessard, N., Paré, M., Lepore, F., & Lassonde, M. (1998). Early blind human subjects localize sound sources better than sighted subjects. *Nature, 395*, 278–280.

Macmillan, N., & Creelman, D. (1991). *Detection theory: A user's guide.* New York: Cambridge University Press.

Maguinness, C., & von Kriegstein, K. (2017). Cross-modal processing of voices and faces in developmental prosopagnosia and developmental phonagnosia. *Visual Cognition*. http://dx.doi.org/10.1080/13506285.2017.1313347.

Moro, S. S., & Steeves, J. K. E. (2012). No Colavita effect: Equal auditory and visual processing in people with one eye. *Experimental Brain Research, 216*, 367–373.

Moro, S. S., Harris, L. R., & Steeves, J. K. E. (2014). Optimal audiovisual processing in people with one eye. *Multisensory Research, 27*, 173–188.

Moro, S. S., Kelly, K. R., McKetton, L., Gallie, B. L., & Steeves, J. K. E. (2015). Evidence of multisensory plasticity: Asymmetrical medial geniculate body in people with one eye. *Neuroimage: Clinical, 9*, 513–518.

Reed, M. J., Steeves, J. K. E., Kraft, S. P., Gallie, B. L., & Steinbach, M. J. (1996). Contrast letter thresholds in the non-affected eye of strabismic and unilateral eye enucleated children. *Vision Research, 36*, 3011–3018.

Röder, B., Stock, O., Bien, S., Neville, H., & Rösler, F. (2002). Speech processing activates visual cortex in congenitally blind humans. *European Journal of Neuroscience, 16*, 930–936.

Röder, B., Teder-Sälejärvi, W., Sterr, A., Rösler, F., Hillyard, S. A., & Neville, H. J. (1999). Improved auditory spatial tuning in blind humans. *Nature, 400*, 162–166.

Sathian, K. (2000). Practice makes perfect: Sharper tactile perception in the blind. *American Academy of Neurology, 54*, 2203.

Schweinberger, S. R., Herholz, A., & Sommer, W. (1997). Recognizing famous voices: Influence of stimulus duration and different types of retrieval cues. *Journal of Speech, Language and Hearing Research, 40*, 453–463.

Sheffert, S. M., & Olson, E. (2004). Audiovisual speech facilitates voice learning. *Perception and Psychophysics, 66*, 352–362.

Steeves, J. K. E., Wilkinson, F., González, E. G., Wilson, H. R., & Steinbach, M. J. (2004). Global shape discrimination at reduced contrast in enucleated observers. *Vision Research, 44*, 943–949.

von Kriegstein, K., & Giraud, A. (2006). Implicit multisensory associations influence voice recognition. *PLoS Biology, 4*, 1809–1820.

von Kriegstein, K., Dogan, O., Gruter, M., Giraud, A. L., Kell, C. A., Gruter, T., et al. (2008). Simulation of talking faces in the human brain improves auditory speech recognition. *Proceedings of the National Academy of Sciences of the United States of America, 105*, 6747–6752.

Welch, R. B., & Warren, D. H. (1980). Immediate perceptual response to intersensory discrepancy. *Psychological Bulletin, 88*, 638–667.