Journal section: Cognitive Neuroscience

Brain responses to audiovisual speech mismatch in infants are associated with individual differences in looking behaviour.

Kushnerenko, Elena[1]*; Tomalski, Przemyslaw[1,2]*; Ballieux, Haiko[1]; Ribeiro, Helena[1]; Potton, Anita[1]; Axelsson, Emma L.[1]; Murphy, Elizabeth[1] and Moore, Derek G[1].

[1] Institute for Research in Child Development, School of Psychology, University of East London, London, UK

[2] Faculty of Psychology, University of Warsaw, Warsaw, Poland

* EK and PT contributed equally to this work.

Corresponding author:

Elena Kushnerenko, PhD, Institute for Research in Child Development, School of Psychology, University of East London, Water Lane, London E15 4LZ, UK

Phone: +44 (0) 79 13 88 73 05, Fax: +44 (0) 20 8223 4937, E-mail:

e.kushnerenko@gmail.com

Running title: brain responses to audiovisual speech in infants

The total number of pages – 21 pages
Figures: 2
Tables: 2
The total number of words in: the whole manuscript – 5,408

in Abstract – 230 words

in Introduction – 598 words

Keywords: event-related potentials (ERPs), audiovisual mismatch response, McGurk illusion, eye-tracking, audiovisual integration

Abstract

Research on audiovisual speech integration has reported high levels of individual variability, especially among young infants. In the present study we tested the hypothesis that this variability results from individual differences in the maturation of audiovisual speech processing during infancy. A developmental shift in selective attention to audiovisual speech has been demonstrated between 6 and 9 months with an increase in the time spent looking to articulating mouths as compared to eyes (Lewkowicz & Hansen-Tift, 2012; Tomalski *et al.*, 2012). In the present study we tested whether these changes in behavioural maturational level are associated with differences in brain responses to audiovisual speech across this age range. We measured high-density event-related potentials (ERPs) in response to videos of audio-visually matching and mismatched syllables /ba/ and /ga/; and subsequently examined visual scanning of the same stimuli with eye-tracking. There were no clear age-specific changes in ERPs, but the amplitude of audiovisual mismatch response (AVMMR) to the combination of visual /ba/ and auditory /ga/ was strongly negatively associated with looking time to the mouth in the same condition. These results have significant implications for our understanding of individual differences in neural signatures of audiovisual speech processing in infants, suggesting that they are not strictly related to chronological age but instead associated with the maturation of looking behaviour, and develop at individual rates in the second half of the first year of life.

Abbreviations: AV – audiovisual, AVMMR – audiovisual mismatch response, ERP – event-related potential, ET – eye-tracking

1. Introduction

Audiovisual speech integration as demonstrated originally by McGurk and MacDonald (McGurk & MacDonald, 1976) is a phenomenon where seeing non-matching lip articulation interferes with the perception of a speech sound. In this study, two types of speech illusions were observed, a "fusion" - where visual /ga/ dubbed onto auditory /ga/ (VgaAba) was perceived as /da/, and a "combination" - where a visual /ba/ dubbed onto auditory /ga/ was perceived as /bga/.

Subsequent studies have indicated that infants also may perceive VgaAba stimuli as a 'fusion' (Rosenblum et al., 1997; Burnham & Dodd, 2004; but see Desjardins & Werker, 2004). Less investigated in infancy is the 'combination' condition.

Recent evidence from electrophysiological studies suggests that infants as young as 5-months-old differently process these two types of audio-visually incongruent stimuli, that lead to 'combination' and 'fusion' effects.

In a study by Kushnerenko and colleagues (2008) an audiovisual mismatch response (AVMMR) was found in response to the VbaAga-combination, but not for a VgaAba-fusion. The authors concluded that, whereas an audiovisual mismatch is detected by infants for the VbaAga-combination, this is not perceived for the VgaAba-fusion. They suggested this is because the integration of VgaAba might be 'easier' to achieve. Since open-mouthed visual /ga/ is less predictive (Van Wassenhove *et al.*, 2005) and therefore less in conflict with the auditory stimulus, the VgaAba condition is easier to combine into a fused percept. By contrast the visual /ba/ in the VbaAga-combination, is more predictive due to a specific lip movements and may lead to a perception of a cross-modal mismatch.

Recent eye-tracking data corroborates this interpretation: 6-month-old infants discriminated between the VgaAba-fusion and the VbaAga-combination in terms of

the duration of looking to the mouth (Tomalski *et al.*, 2012). They looked significantly shorter in the VbaAga-combination than either in the VgaAba-fusion or canonical /ba/ and /ga/ conditions.

The role of visual attention in audiovisual integration remains a matter of debate. One line of research suggests that high attentional load or a distracter moving across a speaking face (but not obscuring the mouth) could affect the quality of audiovisual integration in adults (Tiippana et al., 2004; Alsius et al., 2005; Schwartz, 2010), while other studies indicate that even in the absence of directed attention to the mouth it is not possible to completely ignore mouth movements (Buchan & Munhall, 2011, Paré et al., 2003). However, the processes that are almost automatic in adults may require more attention resources in infants. Of particular interest is the developmental shift in selective visual attention to speaking faces during the first year of life: while younger infants attend predominantly to the eyes, this preference changes to increased looking to the mouth during the second half of the first year (Lewkowicz & Hansen-Tift, 2012; Tomalski *et al.*, 2012; Wagner *et al.*, 2013). By the age of 9 months looking to the mouth for VbaAga-combination increases significantly so that there is no longer any difference in looking times during presentation of these two types of incongruent stimuli: the combination and the fusion (Tomalski et al., 2012).

In the present study, we asked whether the increased looking time to the mouth between 6 and 9 months of age indicates either: 1) an increased interest in audiovisual mismatch or 2) an enhanced use of visual speech cues in an attempt to integrate the auditory and visual information. In the first scenario, the amplitude of the AVMMR would be expected to increase due to enhanced processing of the mismatched information, while the opposite pattern could be expected if the audiovisual cues are perceived to be integrated.

2. Methods

We employed the same stimuli used in Kushnerenko et al. (2008) and Tomalski et al. (2012), that is, audio-visually matching and mismatching videos of female faces pronouncing /ba/ and /ga/ syllables; and used both electrophysiological (ERP) and eye-tracking (ET) techniques, with all infants assessed within one testing session. This allowed us to obtain real-time indices of audiovisual information processing and detailed direct measures of visual face scanning within a short time frame. The ERP recordings were always performed before the eye-tracking sessions so that the infants would not become familiar with the AV stimuli prior to ERP testing, thus minimising habituation of neural responses. A separate eye-tracking-only control study confirmed that there was no effect of the order of presentation on eye-tracking results (see Supporting Information, Control study 1).

2.1 Participants.

Twenty-two healthy, full-term infants (6 boys) aged between 6 and 9 months (mean age 30.7 weeks, SD = 4.3 weeks) took part in both the eye-tracking (ET) and ERP tasks. The study was approved by the University of East London Ethics Committee and conformed with the Code of Ethics of the World Medical Association (Declaration of Helsinki). Parents gave written informed consent for their child's participation prior to the study.

2.2. Stimuli.

Video clips were recorded with three female native English speakers articulating /ba/ and /ga/ syllables. Sound onset was adjusted in each clip to 360 ms from stimulus onset, and the auditory syllables lasted for 280 – 320 ms. Video clips were rendered

with a digitization rate of 25 fps, and the stereo soundtracks were digitized at 44.1 kHz with a 16-bit resolution.

The total duration of all AV stimuli was 760 ms. Lips movements started ~260-280 ms before the sound onset (for all speakers). Each AV stimulus started with lips fully closed and was followed immediately with the next AV stimulus, the stimulus onset asynchrony (SOA) being 760 ms, thus giving an impression of a continuous stream of sounds being pronounced. The paradigm was designed as a continuous speech flow specifically to minimize the input of face- and movement- related visual evoked potentials. In order to examine how much of the ERP amplitude is explained by the visual evoked potentials (VEP), an additional control study was carried out with auditory stimuli only (see Supporting information, Control study 2, Figure S1).

For each of the three speakers, four categories of AV stimuli were created: congruent visual /ba/ – auditory /ba/ (VbaAba), visual /ga/ - auditory /ga/ (VgaAga), and two incongruent pairs. The incongruent pairs were created from the original AV stimuli by dubbing the auditory /ba/ onto a visual /ga/ (VgaAba-fusion) and vice versa (VbaAga-combination). Therefore, each auditory and each visual syllable was presented with equal probability and frequency during the task. For more information on the stimuli see Kushnerenko et al. (2008).

2.3. ERP task procedure.

The syllables were presented in a pseudorandom order, with speakers being changed approximately every 40 s to maintain the infants' attention. Videos were displayed on a CRT monitor (30 cm diameter, 60 Hz refresh rate) with a black background while the infant, sitting on a parent's lap, watched them from an 80cm distance in an

acoustically and electrically shielded booth. The faces on the monitor were approximately life size at that distance. Sounds were presented through two loudspeakers behind the screen at about a 65 dB level. The recording time varied from 4 to 6 minutes, depending on each infant's attention to the stimuli. The behaviour of the infants was videotaped and off-line coded for EEG artefact rejection.

2.4. ERP recording and analysis.

High-density EEG was recorded using a 128-channel Hydrocel Sensor Net (EGI Inc.) referenced to the vertex (Tucker, 1993). The EEG signal was amplified, digitized at 500 Hz, and band-pass filtered from 0.1 to 200 Hz. The signal was off-line low-pass filtered at 30 Hz and segmented into epochs starting 100 ms before and ending 1,000 ms after the AV stimulus onset. Channels contaminated by eye or motion artefacts were rejected manually, and trials with more than 20 bad channels were excluded. In addition, video recordings of the infants' behaviour were coded frame-by-frame, and trials during which the infant did not attend to the face were excluded from further analysis. Following artefact rejection, the average number of trials for an individual infant accepted for further analysis was 37.4 for /ba/, 36.7 for /ga/, 37.6 for VgaAba, and 37.8 for VbaAga. Although uncommon for adult ERP studies, this number of accepted trials has been proved to be sufficient in infant studies (Bristow et al., 2009; Dehaene-Lambertz & Dehaene, 1994; Friederici, Friedrich, & Christophe, 2007; Guiraud et al., 2011; Kushnerenko et al., 2008).

Artefact-free segments were re-referenced to the average reference and then averaged for each infant within each condition. A baseline correction was performed by subtracting mean amplitudes in the 260-360 ms window from the video onset (i.e. immediately before the sound onset) to minimise the effects of any ongoing

processing from the preceding stimulus.

According to Kushnerenko et al. (2008) the AVMMR resembled the auditory mismatch response (MMR) and was observed mainly over right fronto-central area (between F4, C4 and Cz), commencing at ~290 ms from the sound onset and lasting beyond the epoch of analysis. In this report AVMMR was observed only in response to apparent audio-visual mismatch of speech cues (visual /ba/ auditory /ga/).

In order to link individual differences in electrophysiological mismatch response to the development of visual scanning the mean amplitude between 290 and 390 ms after sound onset (650-750 ms from video onset) from the area between F4, C4 and Cz was entered into hierarchical linear regression as the dependent variable with looking times to articulating mouth and control demographic variables (age, gender, and second language experience[1]) as predictors.

For the comparison between age groups we also measured mean voltage between 140 to 240 ms from the sound onset, centred around the mean latency of the auditory infantile P2 (Kushnerenko *et al.*, 2002a, 2007) over the frontal leads. Repeated-measures ANOVAs were performed on average amplitudes across channels within the following channel groups: frontal (area between Fp1, F3 and Fz on the left and symmetrical on the right), central (area between F3, C3 and Cz on the left and symmetrical on the right), occipital (area between O1, P3 and Pz on the left and symmetrical on the right) and temporo-parietal (covering area between P3 and left mastoid and P4 and the right mastoid, Figure S2). The Greenhouse–Geisser correction was used where necessary.

For illustration purposes topographic maps of the voltage difference between stimuli

[1] Second language experience here is defined as experience of one or more languages at home in addition to the English language.

with the same visual input (that is, VbaAga-minus-VbaAba) were created to eliminate the possible contribution of the visual input.

2.5. Eye-tracking (ET) task

The ET task was presented immediately after the ERP task. Each trial contained 10 repetitions of one instance of the same stimuli used in the ERP session (that is canonical /ba/ or /ga/ and both crossed stimuli) and was 7600 ms long (760 ms x 10). Participants were seated on a parent's lap in a dimly lit room in front of a Tobii T120 eye-tracker monitor (17" diameter, screen refresh rate 60 Hz, ET sampling rate of 120 Hz, spatial accuracy 0.5 degrees), at a distance of approximately 60 cm. Each infant was calibrated using a five-point routine prior to the experiment to ensure positional validity of gaze measurements (successful calibration of all five points was required). At least 50% of samples were recorded from each infant during each trial. The parent's view of the stimulus monitor was obscured, to prevent any interference with the infant's looking behaviour. Eye movements were monitored continuously during each recording. Each participant observed a total of ten trials. Before each trial, the participant's attention was attracted to the screen by colourful animations with sound, which were terminated as soon as the infant fixated them. The first two and the last two trials were the canonical VbaAba and VgaAga trials, and their order of presentation was counterbalanced for participants. In between them, two instances of the crossed stimuli and the silent face still images were displayed in a random order. Previous studies showed no order effects on infant looking times to the stimuli (Tomalski *et al.*, 2012). The entire sequence lasted approximately 2 minutes.

2.6. ET analysis

The eye-tracking data was analysed within specific Areas-Of-Interest (AOIs): mouth,

eyes and the entire face oval (excluding the hair region and ears, Figure S3). The total looking times (fixation lengths) were calculated off-line for each participant, condition and AOI using the Tobii Studio software package and the Tobii fixation filter (Tobii Inc.). The proportion of fixation durations on the mouth area and the eyes area compared to total looking on the entire face oval was calculated.

3. Results

To date, the audiovisual mismatch response (AVMMR) has not been observed in response to all incongruent AV pairs, but only to the VbaAga-combination condition (Kushnerenko *et al.*, 2008). This was the case in the current study as well, hence in the following results we only describe the combination-VbaAga condition.

Relationship between AVMMR amplitude and looking time to the mouth

First, we have tested the association between attention to the mouth and the AVMMR amplitude using hierarchical linear regression analysis with looking times to articulating mouth as predictors and as the dependent variable the mean voltage of audiovisual mismatch response (AVMMR) over the right fronto-central channels (290-390 ms from sound onset, see Kushnerenko et al., 2008). Demographic variables (age, gender, and second language experience; see Table 1) were entered at the first stage for control purposes only, and they did not predict any variance in AVMMR (R^2 = .011, R^2adj = 0; F(3, 18) = 0.07, p = .976). The variables that represent the looking time at the mouth during four speech ET conditions were entered at the second stage, and these predicted a significant proportion of variance (R^2change = .610; F(4, 14) = 5.65, p = .006). The final model was also significant (R^2 = .622, R^2adj = .433; F(7, 14) = 3.29, p = .028).

Within the final model, only the looking time to the mouth during the VbaAga-combination was significant, showing that it alone predicted unique variance additional to the other looking times (beta = -.784, p = .028).

These results demonstrated a strong association between the time spent looking at the mouth during VbaAga-combination condition and the amplitude of the AVMMR in response to the same stimuli (see Figure 1).

____Figure 1 here____

____Table 1 here____

Individual differences in attention to the mouth and the AVMMR

For illustration purposes, the participants were split into two groups (see Table 2) by their looking preferences (percentage of time spent looking at the mouth while watching the incongruent VbaAga stimuli). Ten infants who spent > 50% of the total face scanning time fixating the mouth in the VbaAga condition, also looked significantly longer to the mouth in all other conditions (two-way ANOVA, main effect of group: $F(1,20)=12.91$, $p=.002$, $\eta^2=.39$). They were assigned to the Mouth-Preference group (MP, average looking time to the mouth in all conditions 67.13% (SD=15.2), Table 2). The remaining 12 infants were assigned to the No-Mouth-Preference group (NMP, average looking time to the mouth in all conditions 38.9% (SD=20.6)).

The AVMMR was only observed in the NMP group but not in the MP group. In the former, the AVMMR was clearly observed at the group level as a prolonged right fronto-central positivity (Figure 2, for more channels see Figures S4 and S5).

Table 2 here

[Figure 2 here]

Testing the effect of age on the AVMMR amplitude

Although there was no significant association of the AVMMR amplitude with age in our regression model, for control purposes infants were split into the younger (6 to 7.5 months, n = 11) and the older group (7.5 to 9 months, n = 11) by median age (see Figure S6). No difference in ERP responses to incongruent AV stimuli was found between the age groups in either time window (no effect of age; 140-240ms: $F(1,20)=.11$, $p=.74$; 290-390ms: $F(1,20)=2.7$, $p=.12$; no age x condition interaction: 140-240ms: $F(1,20)=.66$, $p=.42$; 290-390ms: $F(1,20)=1.29$, $p=.27$).

The results of the present study therefore demonstrate that the presence of specific neural activity in response to the VbaAga-combination in 6- to 9-month-old infants is strongly negatively associated with infants' attention to mouth articulations and cannot be explained by chronological age or other control demographic variables.

4. Discussion

The purpose of this study was to establish whether individual differences in the amount of visual attention to mouth articulations between 6 and 9 months of age are associated with neural signatures of audiovisual speech processing (the ERP audiovisual mismatch response, AVMMR). Given that previous eye-tracking data has shown the presence of developmental change in visual attention to speaking mouth between 6 and 9 months of age (Lewkowicz & Hansen-Tift, 2012; Tomalski *et al.*, 2012), we expected to see a related change in brain responses to audiovisual speech within the same age range.

In particular, we asked whether the increased looking time to the mouth between 6 and 9 months of age indicates either: 1) an increased interest in audiovisual mismatch or 2) an enhanced use of visual speech cues in an attempt to integrate the auditory and visual information. We measured ERPs in response to congruent and incongruent audiovisual speech cues, and subsequently recorded face-scanning patterns using eye-tracking while infants watched the same stimuli. We found a strong association between neural responses (the AVMMR) and the length of looking to the mouth in the same condition (VbaAga-combination). The amplitude of AVMMR (290-390 ms from sound onset) in the ERP task was strongly negatively correlated with looking times to the mouth during the presentation of the VbaAga-combination stimulus in the subsequent eye-tracking task.

The AVMMR is thought to reflect quick automatic brain detection of mismatch between cues from two modalities, similarly to the pre-attentive auditory-only mismatch response (Kushnerenko *et al.*, 2008). Previously it has been shown that the auditory mismatch response (MMR) in infants undergoes a prolonged maturational process with a large positivity gradually decreasing in amplitude between the age of 3 months until approximately the end of the first year of life (Kushnerenko *et al.*, 2002b, Kushnerenko et al, under review; Morr *et al.*, 2002). Moreover, while no group differences were found in auditory ERPs between 6 and 9 months of age, large inter-individual variability was reported (e.g., Kushnerenko, Ceponiene, Balan, Fellman, & Naatanen, 2002), suggesting that this maturational change occurs at different rates in individual infants and is rather loosely related to chronological age (Kushnerenko *et al.*, 2002b).

We suggest that the same principle may be applicable to maturation of audiovisual speech processing. Indeed, in the present study the AVMMR amplitude was

associated with a specific looking preference rather than with chronological age. The AVMMR was only observed in the No-Mouth-Preference (NMP) sub-group, which according to the recent study of Lewkowitz and Hansen-Tift (2012) could be considered less mature in audiovisual processing. In their study older infants have shown looking preference to the mouth while watching speaking faces, while younger infants prefer to look at the eyes. Given this developmental shift the AVMMR may represent a less mature electrophysiological pattern of audiovisual speech processing because it was associated with less time spent looking at the articulatory movements during speech.

The maturational changes in the way auditory and visual information is processed by younger and older infants, are reflected in developmentally transient ERP components, which are reliably elicited in younger infants but are not always observable in older infants and/or adults. For instance, the AVMMR recorded in 2-month-old infants by Bristow and colleagues (Bristow *et al.*, 2009) was not observed in adults (G. Dehaene-Lambertz, personal communication, see also Jääskeläinen et al., 2004), and an increase in visual N290 component to static direct eye-gaze versus averted eye-gaze reported in 4-month-old infants (Farroni *et al.*, 2002) was not observed in 9-month-old infants (Elsabbagh *et al.*, 2009) or adults (Grice *et al.*, 2005). In order to further explore the question of the developmental profile of the AVMMR neural response, a group of adults was also tested (see Supporting information Control study 3 and Figure S7). No AVMMR in response to either audio-visually incongruent (combination and fusion) stimuli was observed[2], confirming our hypothesis that this

[2] **Note that the present study did not employ an oddball paradigm used in previous adult studies** (Saint-Amour *et al.*, 2007; Hessler *et al.*, 2013)**, where AVMMR was elicited in response to the deviant among repetitive standards and not to the audio-visual violation per se. Therefore, the absence of the AVMMR in the present study does not contradict the results of the above studies, but on the contrary provide corroborative evidence that adults perceived both incongruent conditions integrated.**

component indicates a less mature type of processing of AV conflict only in early infancy.

It is not surprising therefore that while the AVMMR was observed at the group level in younger infants (4.5 to 5.5 months, Kushnerenko *et al.*, 2008 and 2 month-old, Bristow et al, 2009), it was only found in the present study in a subset of our infants, who also demonstrated a less mature pattern of looking behaviour. It is important to note here that the group-averaged ERP results might obscure the meaningful individual differences in the level of maturation of multisensory processing in individual infants. Thus, it appears that the AVMMR is a developmentally transient ERP response that may begin to disappear around the age of 6-9 months, similar to mismatch positivity (or PC) in young infants (Morr *et al.*, 2002; Kushnerenko *et al.*, under review). The developmental decrease of the auditory PC during the first year of life was suggested to reflect decreasing sensitivity to less informative sensory cues, which was initially high in younger infants (Kushnerenko et al, under review).

It has been proposed that not integrating cues may be adaptive for young children, as they must learn not only to combine cues but also to establish if these cues are reliable, or if some must be ignored (Nardini *et al.*, 2010). The developmental pattern observed between 6 and 9 months of age in the previous eye-tracking study (Tomalski *et al.*, 2012) is in accordance with this hypothesis: short looking time to the mouth in the mismatched condition indicates that 6-month-old infants try to ignore unreliable and confusing visual cues. Further, the increase of the looking time to the mouth in the same condition by the age of 9 months may indicate the transition from processing of the conflicting cues separately to reducing uncertainty by integrating information.

The absence of the AVMMR in the more behaviourally mature sub-group (MP) of the present study also supports this interpretation: When auditory and visual cues are

perceived as separate, the sensory conflict is detected and the AVMMR is elicited. In the more behaviourally mature group the developing ability to integrate comes at a cost of losing accuracy in the processing of single-cue information and in the ability to detect sensory conflicts (Hillis *et al.*, 2002). A speculation can be made, that with more experience with language and with exposure to different accents or individual pronunciations, multimodal processing may allow better assimilation of inaccurate auditory and visual cues, enabling infants to arrive at the closest possible unified percept. It should be emphasized, though, that this percept might be different for infants and adults.

Therefore, the results of our study have confirmed that the looking times to the mouth in the VbaAga-combination condition were not associated with increased processing of audiovisual mismatch, which should have resulted in an increased amplitude of AVMMR. The results confirmed the second scenario, suggesting that increased looking times to the mouth are associated with the enhanced use of the visual input in an attempt to assimilate ambiguous audiovisual cues to a unified percept. Consequently, as this integration ability strengthens in development, a decreasing (or absent) right-lateralized fronto-central positive AVMMR indicates that sensory conflict is no longer perceived.

The present study demonstrates the importance of combining electrophysiological and behavioural (eye-tracking) measures in identifying the sources of individual variability in infant ERPs. It also suggests that behavioural measures, such as looking preferences, could potentially indicate the level of maturity in the processing and integration of multisensory information.

Acknowledgements

We acknowledge the financial support of Eranda Foundation, and the University of East London (Promising Researcher Grant to EK and School of Psychology funding to PT and DM). We would like to thank Mark Johnson, Gergely Csibra, Kaoru Sekiyama, Jyrki Tuomainen, Robin Panneton, Ghislaine Dehaene-Lambertz and Richard Aslin for helpful discussions, as well as Glorianne Spiteri, and Caroline Frostick for assistance with data collection and Mike Griffiths for advice on some analyses. We thank all families for their participation in the study.

References

Alsius, A., Navarra, J., Campbell, R., & Soto-Faraco, S. (2005) Audiovisual integration of speech falters under high attention demands. *Curr Biol*, 15, 839–843.

Bristow, D., Dehaene-Lambertz, G., Mattout, J., Soares, C., Gliga, T., Baillet, S., & Mangin, J.-F.F. (2009) Hearing faces: how the infant brain matches the face it sees with the speech it hears. *J Cogn Neurosci*, 21, 905–921.

Buchan, J.N. & Munhall, K.G. (2011) The influence of selective attention to auditory and visual speech on the integration of audiovisual speech information. *Perception*, 40, 1164–1182.

Burnham, D. & Dodd, B. (2004) Auditory-visual speech integration by prelinguistic infants: perception of an emergent consonant in the McGurk effect. *Dev Psychobiol*, 45, 204–220.

Dehaene-Lambertz, G. & Dehaene, S. (1994) Speed and cerebral correlates of syllable discrimination in infants. *Nature*, 28, 293–294.

Desjardins, R.N. & Werker, J.F. (2004) Is the integration of heard and seen speech mandatory for infants? *Dev Psychobiol*, 45, 187–203.

Elsabbagh, M., Volein, A., Csibra, G., Holmboe, K., Garwood, H., Tucker, L., Krljes, S., Baron-Cohen, S., Bolton, P., Charman, T., Baird, G., & Johnson, M.H. (2009) Neural correlates of eye gaze processing in the infant broader autism phenotype. *Biological psychiatry*, 65, 31–38.

Farroni, T., Csibra, G., Simion, F., & Johnson, M.H. (2002) Eye contact detection in humans from birth. *Proc Natl Acad Sci U S A*, 99, 9602–9605.

Friederici, A.D., Friedrich, M., & Christophe, A. (2007) Brain responses in 4-month-old infants are already language specific. *Curr Biol*, 17, 1208–1211.

Grice, S.J., Halit, H., Farroni, T., Baron-Cohen, S., Bolton, P., & Johnson, M.H. (2005) Neural correlates of eye-gaze detection in young children with autism. *Cortex*, 41, 342–353.

Guiraud, J.A., Kushnerenko, E., Tomalski, P., Davies, K., Ribeiro, H., Johnson, M.H., & Team, T.B. (2011) Differential habituation to repeated sounds in infants at high risk for autism. *Neuroreport*, 22, 845–849.

Hessler, D., Jonkers, R., Stowe, L., & Bastiaanse, R. (2013) The whole is more than the sum of its parts - Audiovisual processing of phonemes investigated with ERPs. *Brain Lang*, 124, 213–224.

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

Jääskeläinen, I.P., Ojanen, V., Ahveninen, J., Auranen, T., Levänen, S., Möttönen, R., Tarnanen, I., & Sams, M. (2004) Adaptation of neuromagnetic N1 responses to phonetic stimuli by visual speech in humans. *Neuroreport*, 15, 2741–2744.

Kuhl, P.K. & Meltzoff, A.N. (1982) The bimodal perception of speech in infancy. *Science*, 218, 1138–1141.

Kushnerenko, E., Čeponienė, R., Balan, P., Fellman, V., Näätänen, R., & Huotilainen, M. (2002a) Maturation of the auditory event-related potentials during the 1st year of life. *NeuroReport*, 13, 16–22.

Kushnerenko, E., Čeponienė, R., Balan, P., Fellman, V., Näätänen, R., & Huotilainen, M. (2002b) Maturation of the auditory change-detection response in infants: A longitudinal ERP study. *NeuroReport*, 13, 1843–1848.

Kushnerenko, E., Teinonen, T., Volein, A., & Csibra, G. (2008) Electrophysiological evidence of illusory audiovisual speech percept in human infants. *Proc Natl Acad Sci U S A*, 105, 11442–11445.

Kushnerenko, E., Van den Bergh, B.R.H., & Winkler, I. Separating acoustic deviance from novelty during the first year of life: A review of event related potential evidence. *Frontiers in Developmental Pshychology*, under review.

Kushnerenko, E., Winkler, I., Horváth, J., Näätänen, R., Pavlov, I., Fellman, V., & Huotilainen, M. (2007) Processing acoustic change and novelty in newborn infants. *Eur J Neurosci*, 26, 265–274.

Lewkowicz, D.J. & Hansen-Tift, A.M.C.-P. (2012) Infants deploy selective attention to the mouth of a talking face when learning speech. *Proc Natl Acad Sci U S A*, 109, 1431–1436.

McGurk, H. & MacDonald, J. (1976) Hearing lips and seing voices. *Nature*, 264, 746–748.

Morr, M.L., Shafer, V.L., Kreuzer, J.A., & Kurtzberg, D. (2002) Maturation of mismatch negativity in typically developing infants and preschool children. *Ear Hear*, 23, 118–136.

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

Patterson, M.L. & Werker, J.F. (2003) Two-month-old infants match phonetic information in lips and voice. *Dev Sci*, 6, 191–196.

Rosenblum, L.D., Schmuckler, M.A., & Johnson, J.A. (1997) The McGurk effect in infants. *Percept Psychophys*, 59, 347–357.

Ross, L.A., Molholm, S., Blanco, D., Gomez-Ramirez, M., Saint-Amour, D., & Foxe, J.J. (2011) The development of multisensory speech perception continues into the late childhood years. *Eur J Neurosci*, 33, 2329–2337.

Saint-Amour, D., De Sanctis, P., Molholm, S., Ritter, W., & Foxe, J.J.C.-P. (2007) Seeing voices: High-density electrical mapping and source-analysis of the multisensory mismatch negativity evoked during the McGurk illusion. *Neuropsychologia*, 45, 587–597.

Schwartz, J.L. (2010) A reanalysis of McGurk data suggests that audiovisual fusion in speech perception is subject-dependent. *J Acoust Soc Am*, 127, 1584–1594.

Tiippana, K., Andersen, T.S., & Sams, M. (2004) Visual attention modulates audiovisual speech perception. *Eur J of Cogn Psychol*, 16, 457–472.

Tomalski, P., Ribeiro, H., Ballieux, H., Axelsson, E.L., Murphy, E., Moore, D.G., & Kushnerenko, E. (2012) Exploring early developmental changes in face scanning patterns during the perception of audiovisual mismatch of speech cues. *Eur J Dev Psychol*, 1–14.

Trainor, L., McFadden, M., Hodgson, L., Darragh, L., Barlow, J., Matsos, L., & Sonnadara, R. (2003) Changes in auditory cortex and the development of mismatch negativity between 2 and 6 months of age. *Int J Psychophysiol*, 51, 5–15.

Tremblay, C., Champoux, F., Voss, P., Bacon, B.A., Lepore, F., & Théoret, H.C.-P. (2007) Speech and non-speech audio-visual illusions: a developmental study. *PLoS One*, 2, e742.

Tucker, D.M. (1993) Spatial sampling of head electrical fields: the geodesic sensor net. *Electroencephalogr Clin Neurophysiol*, 87, 154–63 ST – Spatial sampling of head electrical f.

Van Wassenhove, V., Grant, K.W., & Poeppel, D. (2005) Visual speech speeds up the neural processing of auditory speech. *Proc Natl Acad Sci U S A*, 102, 1181–1186.

Wagner, J.B., Luyster, R.J., Yim, J.Y., Tager-Flusberg, H., & Nelson, C. a. (2013) The role of early visual attention in social development. *International Journal of Behavioral Development*, 37, 118–124.

Figure 1. A scatterplot showing the relationship between the mean voltage over the central-right area (AVMMR) and looking times to the mouth during the presentation of the incongruent VbaAga stimuli.

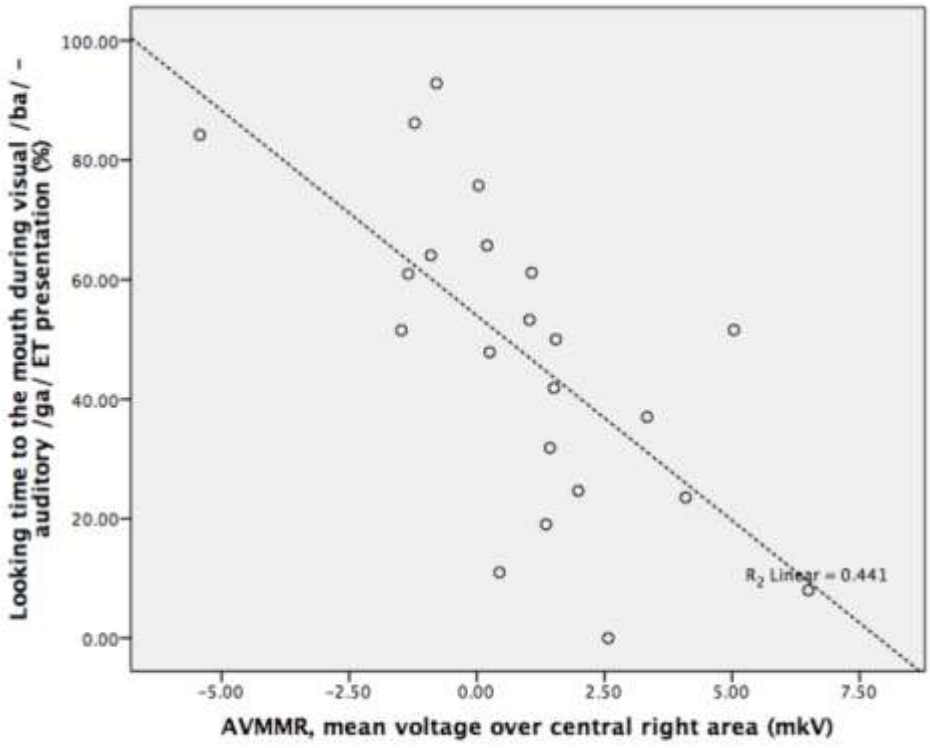

Figure 2. Grand-averaged ERP responses for the No-Mouth-Preference (NMP) group (A) and Mouth-Preference group (B) to the audiovisual stimuli: VbaAba (thin grey), VbaAga (orange), VgaAba (blue) and VgaAga (black, dotted). Topographic maps represents the voltage difference between responses to VbaAga and VbaAba pairs within the time window 290 to 390 ms.