

Attention to speaking mouth is reduced by face inversion in infants only from 9 months of age

Zuzanna Laudańska (✉ zlaudanska@psych.pan.pl)

Institute of Psychology, Polish Academy of Sciences

Aleksandra Dopierała

University of Warsaw

Magdalena Szmytke

University of Warsaw

Dianna Ilyka

University of Cambridge

Anna Malinowska-Korczak

Institute of Psychology, Polish Academy of Sciences

Przemysław Tomalski

University of Warsaw

Research Article

Keywords: Audiovisual speech processing, Face processing, McGurk effect, Face inversion effect, Infants, Eye-tracking

Posted Date: March 24th, 2021

DOI: <https://doi.org/10.21203/rs.3.rs-352367/v1>

License:   This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Attention to speaking mouth is reduced by face inversion in infants only from 9 months of age

Authors:

Zuzanna Laudańska^{1*}, zlaudanska@psych.pan.pl, ORCID: 0000-0001-6790-9559

Aleksandra Dopierała², aleksandra.dopierala@psych.uw.edu.pl, ORCID: 0000-0003-3115-4806

Magdalena Szmytke², m.szmytke2@uw.edu.pl, ORCID: 0000-0003-4714-0601

Dianna Ilyka³, di257@cam.ac.uk

Anna Malinowska-Korczak¹, a.malinowska-korczak@psych.pan.pl, ORCID: 0000-0003-2788-4865

Przemysław Tomalski^{2,1,*}, tomalski@mac.com, ORCID: 0000-0002-0390-5759

¹Neurocognitive Development Lab, Institute of Psychology, Polish Academy of Sciences, Warsaw, Poland

²Neurocognitive Development Lab, Faculty of Psychology, University of Warsaw, Warsaw, Poland

³Department of Psychology, University of Cambridge, Cambridge, the United Kingdom

Abstract

Configural processing is a specialised perceptual mechanism that allows adult humans to quickly process facial information. It emerges before the first birthday and can be disrupted by upside-down presentation of the face (inversion). To date, little is known about the relationship of configural face processing to the emerging knowledge of audiovisual (AV) speech in infancy. Using eye-tracking we measured attention to speaking mouth in upright and inverted faces that were either congruent or incongruent with the speech sound. Face inversion affected looking at AV speech only in older infants (9- to 11- and 12- to 14-month-olds). The youngest group of infants (5- to 7-month-olds) did not show any differences in looking durations between upright and inverted faces, while in both older groups face inversion led to reduced looking at the articulating mouth. We also observed a stronger interest in the eyes in the youngest infants, followed by an increase in looking time to the mouth in both older groups. Our findings suggest that configural face processing is involved in AV speech processing already in infancy, indicating early integration of face and speech processing mechanisms in cognitive development.

Keywords

Audiovisual speech processing, Face processing, McGurk effect, Face inversion effect, Infants, Eye-tracking

Introduction

Learning about speech in infancy is a multimodal process that occurs during early social interactions. From birth, infants pay special attention to faces, which provide communicative information (such as emotional expressions or gaze direction) together with temporally-synchronised audiovisual (AV) speech cues. Speech is a source of cross-modal redundancy [1,2,3], and already at the age of two months infants can match audible vowels with silent visual articulations of native phoneme categories [4]. By 5 months of age, infants can detect AV incongruence and sometimes integrate incongruent visual and auditory information [5,6,7,8, for a review, see 9]. AV speech integration has previously been tested in infants using the McGurk effect [10], which involves perceiving an illusory percept /da/ or /θa/, when observing articulations of /ga/ while hearing an incongruent speech sound /ba/. A reverse combination of visual /ba/ and auditory /ga/ typically leads to audiovisual mismatch perceived as /bga/ or /baga/. These perceptual illusions observed even in early infancy make the McGurk effect a prime candidate for investigating the AV speech perception and integration across development.

While speech perception in adults typically involves the processing of speech sounds, information obtained from faces also affects how speech is perceived and processed. The presence of visual cues such as articulations supports speech comprehension and phonological learning [11,12]. On the other hand, face inversion disrupts the processing of spatial relations among facial features (configural information) and diminishes the AV speech integration as revealed by studies using the McGurk paradigm. Specifically, adults report less illusory percepts when viewing inverted relative to upright faces [13]. Moreover, the mismatch response (MMN), an electrophysiological marker of deviance detection, was observed in response to illusory percepts only for upright but not for inverted faces [14]. The effect of face inversion on AV speech perception, demonstrated in these studies, supports Calvert and Campbell's [15] suggestion that AV speech processing requires configural facial information to deliver specific visible patterns that suggest particular mouth articulation. Rosenblum and colleagues [16] suggested that this relationship between face and audiovisual/visual speech processing reflects the specialisation for upright face processing, raising further questions on whether and how these associations emerge over the course of development. Evidence shows that face and speech processing undergo significant and parallel transitions throughout the first year of infants' lives [17]. Face inversion effect, a marker of configural face processing and hallmark of adult's expertise in upright faces [18,19], is initially observed around 4-5 months of age and undergoes fast-paced developmental changes across the second half of the first year of life [20,21,22]. Moreover, by the age of 10-12 months, infants' ability to process a wide

variety of faces (i.e., other-race and other-species faces) declines to include only those often seen in their environment [17]. Similarly, the acquisition of speech sounds and speech categories is an experience-dependent process following which infants, initially proficient in distinguishing a variety of native and non-native speech contrasts, gradually specialize to process only those sounds that are frequent in their environment [23,24].

Despite striking developmental parallels and ubiquitous co-occurrence of upright and speaking faces, the specialisation of faces and speech has been mostly studied separately. However, recent studies show that even though face and speech perception are initially unrelated to each other, they become positively related by the end of the first year of life. First, Minar and Lewkowicz [25] showed that infants aged 10-12 months maintained their ability to discriminate dynamic other-race faces accompanied by a speech syllable, whereas they did not discriminate between dynamic other-race silent faces or other-race faces accompanied by a non-speech sound. Second, Krasotkina and collaborators [26] found a positive correlation between the ability to discriminate between non-native tones and other-race faces - infants who were weak at discriminating stimuli in one domain were also weak at discriminating stimuli in the other domain. Third, Xiao and collaborators [27] showed that narrowing in the face and narrowing in speech perception were not correlated at or before 6 months of age, but they were negatively correlated at 9 months and positively correlated at 12 months. Finally, Ujiie et al. [28] observed the McGurk effect in 8-to 9-month-old infants watching own-race faces, but not in those watching the other-race faces. They showed that infants could perceive the McGurk effect only when the voice is paired with a face similar to the ones featuring in the infants' social environments. These studies support the view that developmental trajectories for faces and speech perception may share some underlying mechanisms that drive the processes of specialisation in both domains.

Furthermore, the pattern of scanning articulating faces reveals that attention to different facial features (eyes, mouth) is related to the development of speech processing skills. Infants' attention to the mouth increases between 4 and 8 months of age [29,8,30,31; but see also different pattern in 32] as attention to the talking mouth provides infants with important information that may facilitate their phonological development [33,34,35]. Lewkowicz and Hansen-Tift [29] showed that during the AV speech perception task, 4-month-olds attend relatively more to the eyes, 6-month-olds attend equally to the eyes and mouth, 8- and 10-month-olds attend more to the mouth, while 12-month-olds attend equally to the eyes and mouth, and adults attend more to the eyes.

In the present study, we investigated whether configural face processing affects AV speech perception in infancy. We hypothesised that, if the processing of faces and AV speech share some

underlying mechanisms, face inversion will affect infants' scanning patterns of AV speech cues. Using eye-tracking, in infants aged 5-7, 9-11 and 12-14 months we tested visual attention to the eyes and mouth of the upright and inverted faces of speakers articulating syllables with congruent or incongruent auditory tracks. As in other McGurk studies on infants [7,36,37] we used the audiovisually incongruent syllables /ba/ and /ga/ to produce illusory /da/ (vGAaBA) and mismatch /bga/ (vBAaGA) percepts. In line with earlier studies, we had three predictions. Firstly, face inversion was expected to particularly affect attention to the mouth, given that articulating mouth is the most informative in delivering specific visible pattern of speech sounds [15]. Secondly, this inversion effect should be more pronounced in older infants who attend more to the mouth than the eyes when processing speech [29] and have developed more mature face processing relative to younger age groups [17]. Thirdly, we expected a non-linear developmental pattern of attention towards the eyes and the mouth in upright congruent conditions, with an initial greater looking at the eyes area at 5-7-months, followed by increased relative preference in looking at the articulating mouth at 9-11 months of age and more equally distributed attention between the talker's eyes and mouth at 12-14 months of age.

Methods

Participants

A total of 120 infants, all healthy and born full-term (36 gestational weeks or more), contributed data for analysis. Participants were infants from predominantly middle-class families living in the city with >1.5 million inhabitants. Each infant was Polish monolingual and belonged to one of the following age groups: 5-7-month-olds, 9-11-month-olds and 12-14-month-olds. A full breakdown of participants by age group is presented in Table 1. An additional 64 infants were tested but excluded from the final sample because of preterm birth (N = 2), fussiness or failure to reach looking time criteria (N = 46), bilingualism (hearing Polish at home for less than 40% of the time, N = 7), technical difficulties (N = 8) or experimenter error (N = 1). The study was approved by the Research Ethics Committee at the Faculty of Psychology, University of Warsaw, Poland, and conformed with the standards of the Declaration of Helsinki. Prior to the testing, all parents gave written informed consent. For their participation, the families received a diploma and a small gift (a baby book) and a video recording of their play in the laboratory.

Table 1

Participant Characteristics

	5-7-month-olds	9-11-month-olds	12-14-month-olds
N (girls)	39 (17)	41 (19)	40 (15)
Age in months (mean, SD)	5.85 (0.46)	9.91 (0.47)	12.99 (0.65)
Birth weight in grams (mean, SD)	3396 (568)	3511 (639)	3566 (372)

Procedure

The eye-tracking task reported here for 5-7- and 9-11-months-old infants was a part of a longer testing session which included Parent-Child Interaction and questionnaire data collection. The study had a cross-sectional design. The eye-tracking experiment took place in a purpose-built, infant-friendly testing-room. The area around the eye-tracker monitor was covered with a black cloth to provide a uniform background and occlude potential visual distractors. During the eye-tracking task, infants sat in an infant high chair or on parent's lap, approximately 60 cm from the monitor. Eye-tracking data were collected using Tobii T60XL (Tobii AB, Sweden) with a 24"

monitor, recording at a 60 Hz sampling rate. Infant gaze position was calibrated with colorful animation using a 5-point routine (four points in corners of the screen and one in the centre). All participants successfully calibrated at least 4 points (calibrated from 1 to 3 times; $M = 4,69$, $SD = 0,466$). The infant's behavior was monitored throughout the task via camera placed above the screen. The stimuli were presented using Tobii Studio Pro 3.2.2. (Tobii AB, Sweden) running on a HP EliteBook 8470w laptop.

Stimuli

Infants were presented with short videos of two female native Polish speakers articulating /ba/ or /ga/. For each speaker eight categories of audiovisual stimuli were created: (1) upright congruent audiovisual /ba/; (2) upright congruent audiovisual /ga/; (3) upright non-fusible incongruent: visual /ba/ and audio /ga/; (4) upright fusible incongruent: visual /ga/ and audio /ga/; (5) inverted congruent audiovisual /ba/; (6) inverted congruent audiovisual /ga/; (7) inverted non-fusible incongruent: visual /ba/ and audio /ga/; (8) inverted fusible incongruent: visual /ga/ and audio /ga/. The incongruent pairs were created from original AV stimuli by dubbing the auditory /ba/ onto a visual /ga/ and vice versa. Sound onset was adjusted to each video clip to 360 ms from visual stimulus onset. The inverted stimuli were created from both congruent and incongruent AV stimuli by turning them 180 degrees upside-down. Infants viewed 16 trials (two trials per condition) in a single fixed order block. Before each trial, the infant's attention was attracted to the centre of the screen by presentation of a spinning cartoon with sound, which was terminated by the experimenter as soon as the infant focused on it. After 8 trials, an infant-friendly cartoon movie was presented for a maximum of 20s to prevent boredom. The procedure took approximately 4 minutes.

Data analysis

For each trial, the total looking time was extracted using Tobii Studio Pro 3.2.2. in three Areas of Interest (AOI): mouth region, eyes region and the entire face oval (excluding the hair region and ears). Each AOI was intentionally bigger to allow for slight head movements occurring during speech articulation. Data were excluded in trials, where infant looked at the entire face region for less than 3 seconds. Only infants with at least 11 good trials out of 16 were included in the analyses. Infants had on average 14.57 ($SD = 1.54$) included trials, and there were no differences between age groups in the number of included trials ($F(2,118) = .009$; $p = 0.991$). A Proportion of Total Looking Time (PTLT) was calculated for each condition and participant by dividing the total looking time to the eyes and mouth, respectively, by the total looking time to the

face (an Eyes-to-Face and Mouth-to-Face ratios). A PTLT Difference Score was also calculated by subtracting the PTLT score for the mouth from the PTLT for the eyes, such that PTLT difference scores above the zero represented a preference for looking at the eyes, and PTLT difference scores below zero represented a greater interest in the mouth.

Repeated-measures ANOVAs were performed on Mouth-to-Face Ratio and PTLT Difference Scores with 4 Audiovisual Stimuli (vBAaBA, vGAaGA, vBAaGA, vGAaBA) x 2 Face Orientation (Upright, Inverted) as repeated measures factors and x 3 Age Groups (5-7- month-olds, 9-11-month-olds, 12-14-month-olds) as between-subject factor.

Results

Mouth-to-Face Ratio

To check whether face inversion affects attention to the speaker's mouth, a three-way mixed ANOVA on Mouth-to-Face Ratio was conducted. It revealed statistically significant main effects of the Face Orientation $F(1,112) = 17.084$, $p < .001$, $\eta_p^2 = .132$, Stimulus $F(3,336) = 4.346$, $p = .005$, $\eta_p^2 = .037$, Age Group $F(2,112) = 12.53$, $p < .001$, $\eta_p^2 = .183$. There were two statistically significant interactions. The Face Orientation x Age Group interaction $F(2,112) = 3.575$, $p = .031$, $\eta_p^2 = .060$ showed that both older groups of infants looked longer at the mouth of upright than inverted faces ($p = .006$ and $p < .001$, respectively) and 5- to 7-month-olds did not differ significantly in looking at the mouth between upright and inverted conditions. Second interaction of Stimulus x Face Orientation $F(3,336) = 9.293$, $p < .001$, $\eta_p^2 = .077$ indicated that infants looked longer at the mouth of the upright than inverted faces in three speech conditions (vBAaBA, vGAaGA, vBAaGA, all $ps \leq .001$).

To test differences in looking to the mouth between stimuli depending on face orientation we ran a separate ANOVA for each age group. Whereas in the youngest group of infants ($F(3,111) = 1.602$, $p = .193$, $\eta_p^2 = .041$) as well as in 9- to 11-month-olds ($F(3,114) = 1.801$, $p = .151$, $\eta_p^2 = .045$) there was no interaction between the type of stimulus and Face Orientation, it was significant in infants aged 12-14 months, $F(3,111) = 7.938$, $p < .001$, $\eta_p^2 = .177$. In the oldest group, there were significant differences in looking time at the mouth between upright and inverted faces in both congruent conditions as well as in non-fusible incongruent one (vBAaGA; all $ps \leq .001$). However, there was no such difference in for fusible vGAaBA ($p = .168$).

To control whether face inversion was specifically affecting looking to the mouth, a three-way mixed ANOVA on Eyes-to-Face Ratio was conducted. No effects of Face Orientation or Face Orientation x Age Group interaction were found (full description can be found in the Supplementary Information).

Changes in Eyes vs. Mouth Preference

To capture a relative preference in looking at eyes or mouth in different age groups and to check whether the differences between face orientation affect the visual attention, a three-way mixed ANOVA (Stimulus x Face Orientation x Age Group) was conducted on PTLT Difference Scores (see Fig. 1). The analysis showed three main effects: Face Orientation $F(1,110) = 5.161$, $p = .025$, $\eta_p^2 = .045$, infants in upright conditions had lower PTLT difference scores than in the inverted one ($p = .033$); Age Group $F(2,110) = 16.449$, $p < .001$, $\eta_p^2 = .230$, the 5 to 7-month-olds had

significantly higher PTLT difference scores than both older groups (both p s $< .001$); Stimulus $F(3,330) = 5.332$, $p = .001$, $\eta_p^2 = .046$. There was also a significant interaction between Stimulus and Face Orientation, $F(3,333) = 5.580$, $p = .001$, $\eta_p^2 = .048$.

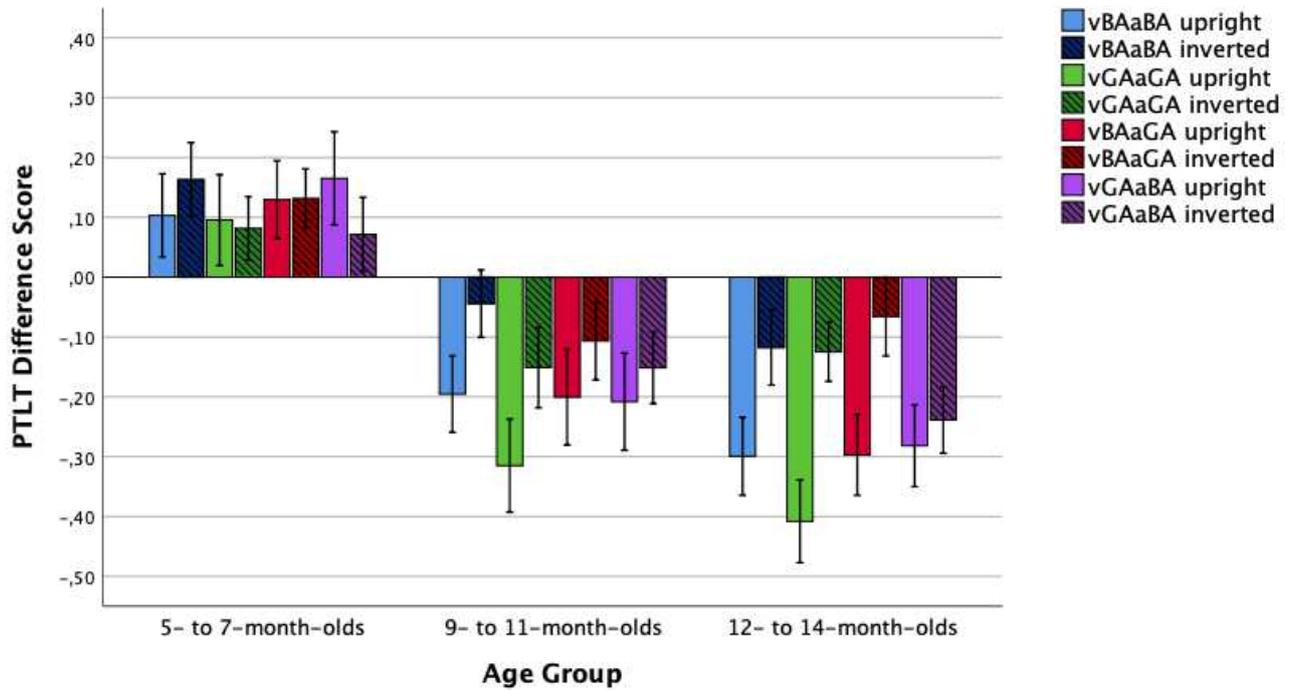


Figure 1. Proportion of Total Looking Time difference scores in each condition and age group. Blue and green bars represent congruent, while red and purple – incongruent AV speech pairs. Error bars represent +/- one standard error of mean.

Discussion

The present study investigated the role of emerging configural face processing in audiovisual speech processing in infants. Using eye-tracking we measured attention to the speaking mouth in upright and inverted faces. Actresses were articulating syllables that were either congruent or incongruent with the speech sound. Supporting our hypothesis, we discovered that face orientation affected the processing of AV speech. This effect was present for older infants whose attention to AV speech cues was particularly affected by inversion. The youngest group of infants (5- to 7-month-olds) did not show any differences in looking durations to the mouth between upright and inverted faces, while in both older groups (9- to 11- and 12- to 14-month-olds) face inversion led to reduced looking to the articulating mouth. We also observed longer looking time to the eyes in 5- to 7 month-old infants, followed by an increase in looking time to the mouth in both older groups. Preferential attention to the speaking mouth was limited to upright but not to inverted faces from the age of 9 months.

Based on adult findings, which showed the role of configural processing in AV speech processing, we expected that the inversion effect would specifically affect infant attention to the speaking mouth. Consistently with that prediction, we found that infants looked longer to the mouth in upright than inverted faces from the age of 9 months. The fact that the effect of inversion was specific to looking towards the mouth region supports the idea that AV speech processing development is related to the emergence of configural face processing. A similar effect was shown by Imafuku and collaborators [38] who found that face inversion makes audiovisual information about vowels more difficult to perceive, which leads to limited imitation. However, their study suggested that inversion affects speech already in 6-month-olds, so earlier than in ours. This discrepancy could be related to the fact that discriminating vowels is observed earlier in development than discriminating consonants, which were used in our experiment [39].

In contrast to previous studies [8], we did not find clear evidence of discrimination between congruent and incongruent regarding looking times to the mouth. However, in our study, the face inversion effect in looking to the mouth region was present in both congruent and mismatch conditions, but not in the incongruent fusible one. The oldest infants looked longer to the mouth in upright than inverted faces in every audiovisual condition except for incongruent fusible (vGAaBA), which suggests that this illusory percept is processed differently than the other AV conditions. We speculate that the attention to the mouth in this condition could result from the effort to maintain AV illusion - even when the face is inverted.

The observation that face inversion affects specifically looking to the articulating mouth region suggests that upright face orientation is especially important for speech perception. Furthermore, we found that older but not younger infants showed a preference for looking at the mouth. When viewing upright faces, 5-7-months-old infants looked relatively longer to the eyes than to the mouth articulating syllables, whereas older infants, 9-11- and 12-14-month-olds, looked relatively longer to the mouth than to the eyes. Our findings are consistent with previous research showing increase in looking to the mouth of a talking face between 4 and 8 months of age [29,8,30,31]. Thus, results from the Polish population are in line with those from English- and French-speaking populations. We haven't observed the second shift in infants' attention postulated by Lewkowicz & Hansen-Tift [29] - the oldest group of infants preferred looking at the mouth in comparison with the eyes. This discrepancy can be potentially explained by the fact that Polish infants are likely to later gain proficiency in consonant discrimination. Vocabulary acquisition is slower in Polish in comparison to English children [40], thus a similar delay may be occurring in the acquisition of phonological knowledge. If so, we could expect infants to continue showing greater interest in the mouth region relative to the eyes for a prolonged period in comparison to English-speaking populations and thus observe the visual scanning pattern similar to the one reported for non-native condition in the study by Lewkowicz & Hansen-Tift [29]. Notwithstanding, this inconsistency could also be related to differences in the stimuli - hearing native and non-native speech monologue (as in [29]) vs. hearing congruent and incongruent syllables and looking at faces presented in upright and inverted orientation (in our experiment). Moreover, both tasks possibly differ in difficulty, which could also influence the results. Nonetheless, the increased attention to the mouth at older age was also described by Morin-Lessard and collaborators [31], who showed that both monolingual (learning English or French) and bilingual (learning both English and French) children from 9 months up to 5 years of age are most interested in the mouth region of the talking faces.

This is the first study to investigate the developmental changes in the relation between configural face and AV speech processing in infants. Our findings indicate that configural face processing is involved in AV speech processing already in the second half of the first year of life. We observed a shift from the eye preference to the mouth preference in the same age point as the emergence of the face inversion effect. Furthermore, as the effect of inversion was visible in looking towards the mouth region, it suggests that upright face orientation is especially important for speech perception. In line with previous studies [25,26,27,28,38], our results support the notion that both the specialisation for speech and the specialisation for faces are related in infancy.

Moreover, the perception of speech and faces may share some underlying neural mechanisms that develop across infancy to achieve efficient processing of talking faces that are most common in infants' social environment.

Author contributions

ZL contributed to the stimulus preparation, collection and analysis of the data and drafting the manuscript. AD contributed to data collection, drafting and editing the manuscript. MS contributed to data analysis, drafting and editing the manuscript. DI contributed to the stimulus preparation, data collection and editing the manuscript. AMK contributed to the stimulus preparation and editing the manuscript, and oversaw data collection. PT conceived and designed the experiment, secured funding, oversaw its execution and contributed to analysis and interpretation of results, edited the manuscript.

Acknowledgements

We thank all families and infants that participated in this study for their generous contribution. The Authors wish to thank Bogumił Karbowski for assistance with stimulus preparation and recruitment. This study was funded by a grant from the National Science Centre of Poland to PT (2016/23/B/HS6/03860). Additional support for data collection in the oldest group of infants was provided by the Institute of Psychology, PAS.

Competing interests

The authors declare no competing interests.

Data availability

The datasets generated during an analysed during the current study are available in the OSF repository, https://osf.io/kvw9n/?view_only=c58c709a63d742dfb9dc194d0d3dc747.

References

1. Bahrick, L., Flom, R. & Lickliter, R. Intersensory redundancy facilitates discrimination of tempo in 3-month-old infants. *Dev. Psychobiol.* **41**, 352-363, <https://doi.org/10.1002/dev.10049> (2002).
2. Bahrick, L., Hernandez-Reif, M. & Flom, R. The Development of Infant Learning About Specific Face-Voice Relations. *Dev. Psychol.* **41**, 541-552, <https://doi.org/10.1037/0012-1649.41.3.541> (2005).
3. Lewkowicz, D. Infant perception of audio-visual speech synchrony. *Dev. Psychol.* **46**, 66-77, <https://doi.org/10.1037/a0015579> (2010).
4. Patterson, M. L., & Werker, J. F. Two-month-old infants match phonetic information in lips and voice. *Dev. Sci.* **6**, 191–196, <https://doi.org/10.1111/1467-7687.00271> (2003).
5. Burnham, D., & Dodd, B. Auditory-visual speech integration by prelinguistic infants: perception of an emergent consonant in the McGurk effect. *Dev. Psychobiol.* **45**, 204-220, <https://doi.org/10.1002/dev.20032> (2004).
6. Desjardins, R. N., & Werker, J. F. Is the integration of heard and seen speech mandatory for infants? *Dev. Psychobiol.* **45**, 187–203, <https://doi.org/10.1002/dev.20033> (2004).
7. Kushnerenko, E., Teinonen, T., Volein, A., & Csibra, G. Electrophysiological evidence of illusory audiovisual speech percept in human infants. *Proc. Natl. Acad. Sci. U. S. A.* **105**, 11442–11445, <https://doi.org/10.1073/pnas.0804275105> (2008).
8. Tomalski, P. *et al.* Exploring early developmental changes in face scanning patterns during the perception of audiovisual mismatch of speech cues. *Eur. J. Dev. Psychol.* **10**, 611-624, <https://doi.org/10.1080/17405629.2012.728076> (2013).
9. Tomalski, P. Developmental Trajectory of Audiovisual Speech Integration in Early Infancy. A Review of Studies Using the McGurk Paradigm. *Psychol. Lang. Commun.* **19**, 77–100, <https://doi.org/10.1515/plc-2015-0006> (2015).
10. McGurk, H., & MacDonald, J. Hearing lips and seeing voices. *Nature* **264**, 746–748, <https://doi.org/10.1038/264746a0> (1976).
11. Hazan, V., Sennema, A., Iba, M., & Faulkner, A. Effect of audiovisual perceptual training on the perception and production of consonants by Japanese learners of English. *Speech Commun.* **47**, 360–378, <https://doi.org/10.1016/j.specom.2005.04.007> (2005).
12. Kawase, T. *et al.* Bimodal audio-visual training enhances auditory adaptation process. *NeuroReport* **20**, 1231–1234, <https://doi.org/10.1097/WNR.0b013e32832fbef8> (2009).
13. Hietanen, J. K., Manninen, P., Sams, M., & Surakka, V. Does audiovisual speech perception use information about facial configuration? *Eur. J. Cogn. Psychol.* **13**, 395–407, <https://doi.org/10.1080/09541440126006> (2001).

14. Eskelund, K., MacDonald, E. N., & Andersen, T. S. Face configuration affects speech perception: Evidence from a McGurk mismatch negativity study. *Neuropsychologia* **66**, 48–54, <https://doi.org/10.1016/j.neuropsychologia.2014.10.021> (2015).
15. Calvert, G. A., & Campbell, R. Reading speech from still and moving faces: The neural substrates of visible speech. *J. Cogn. Neurosci.* **15**, 57–70, <https://doi.org/10.1162/089892903321107828> (2003).
16. Rosenblum, L. D., Yakel, D. A., & Green, K. P. Face and mouth inversion effects on visual and audiovisual speech perception. *J. Exp. Psychol. Hum. Percept. Perform.* **26**, 806–819, <https://doi.org/10.1037/0096-1523.26.2.806> (2000).
17. Maurer, D., & Werker, J. F. Perceptual narrowing during infancy: A comparison of language and faces. *Dev. Psychobiol.* **56**, 154–178, <https://doi.org/10.1002/dev.21177> (2014).
18. Maurer, D., Le Grand, R., & Mondloch, C. J. The many faces of configural processing. *Trends Cogn. Sci.* **6**, 255–260, [https://doi.org/10.1016/S1364-6613\(02\)01903-4](https://doi.org/10.1016/S1364-6613(02)01903-4) (2002).
19. Rossion, B., & Gauthier, I. How does the brain process upright and inverted faces? *Behav. Cogn. Neurosci. Rev.* **1**, 63–75, <https://doi.org/10.1177/1534582302001001004> (2002).
20. Rose, S. A., Jankowski, J. J., & Feldman, J. F. The inversion effect in infancy: The role of internal and external features. *Infant Behav. Dev.* **31**, 470–480, <https://doi.org/10.1016/j.infbeh.2007.12.015> (2008).
21. Kato, M., & Konishi, Y. Where and how infants look: The development of scan paths and fixations in face perception. *Infant Behav. Dev.* **36**, 32–41, <https://doi.org/10.1016/j.infbeh.2012.10.005> (2013).
22. Oakes, L. M., & Ellis, A. E. An Eye-Tracking Investigation of Developmental Changes in Infants' Exploration of Upright and Inverted Human Faces. *Infancy* **18**, 134–148, <https://doi.org/10.1111/j.1532-7078.2011.00107.x> (2013).
23. Werker, J. F., & Tees, R. C. Cross-language speech perception: Evidence for perceptual reorganization during the first year of life. *Infant Behav. Dev.* **7**, 49–63, [https://doi.org/10.1016/S0163-6383\(84\)80022-3](https://doi.org/10.1016/S0163-6383(84)80022-3) (1984).
24. Pascalis, O., De Haan, M., & Nelson, C. A. Is face processing species-specific during the first year of life? *Science* **296**, 1321–1323, <https://doi.org/10.1126/science.1070223> (2002).
25. Minar, N. J., & Lewkowicz, D. J. Overcoming the other-race effect in infancy with multisensory redundancy: 10–12-month-olds discriminate dynamic other-race faces producing speech. *Dev. Sci.* **21**, 1–12, <https://doi.org/10.1111/desc.12604> (2018).
26. Krasotkina, A., Götz, A., Höhle, B., & Schwarzer, G. Perceptual Narrowing in Speech and Face Recognition: Evidence for Intra-individual Cross-Domain Relations. *Front. Psychol.* **9**, 1711, <https://doi.org/10.3389/fpsyg.2018.01711> (2018).
27. Xiao, N. G. *et al.* Narrowing in face and speech perception in infancy: Developmental change in the relations between domains. *J. Exp. Child Psychol.* **176**, 113–127, <https://doi.org/10.1016/J.JECP.2018.06.007> (2018).

28. Ujiie, Y., Asai, T., & Wakabayashi, A. Individual differences and the effect of face configuration information in the McGurk effect. *Exp. Brain Res.* **236**, 973–984, <https://doi.org/10.1007/s00221-018-5188-4> (2018).
29. Lewkowicz, D. J., & Hansen-Tift, A. M. Infants deploy selective attention to the mouth of a talking face when learning speech. *Proc. Natl. Acad. Sci.* **109**, 1431–1436, <https://doi.org/10.1073/pnas.1114783109> (2012).
30. Mercure, E. *et al.* Language experience influences audiovisual speech integration in unimodal and bimodal bilingual infants. *Dev. Sci.* **22**, 1–9, <https://doi.org/10.1111/desc.12701> (2019).
31. Morin-Lessard, E., Poulin-Dubois, D., Segalowitz, N., & Byers-Heinlein, K. Selective attention to the mouth of talking faces in monolinguals and bilinguals aged 5 months to 5 years. *Dev. Psychol.* **55**, 1640–1655, <https://doi.org/10.1037/dev0000750> (2019).
32. Wilcox, T., Stubbs, J. A., Wheeler, L., & Alexander, G. M. Infants' scanning of dynamic faces during the first year. *Infant Behav. Dev.* **36**, 513–516, <https://doi.org/10.1016/j.infbeh.2013.05.001> (2013).
33. Young, G. S., Merin, N., Rogers, S. J., & Ozonoff, S. Gaze behavior and affect at 6 months: Predicting clinical outcomes and language development in typically developing infants and infants at risk for autism. *Dev. Sci.* **12**, 798–814, <https://doi.org/10.1111/j.1467-7687.2009.00833.x> (2009).
34. Tenenbaum, E. J., Sobel, D. M., Sheinkopf, S. J., Malle, B. F., & Morgan, J. L. Attention to the mouth and gaze following in infancy predict language development. *J. Child Lang.* **42**, 1173–1190, <https://doi.org/10.1017/S0305000914000725> (2015).
35. Tsang, T., Atagi, N., & Johnson, S. P. Selective attention to the mouth is associated with expressive language skills in monolingual and bilingual infants. *J. Exp. Child Psychol.* **169**, 93–109, <https://doi.org/10.1016/j.jecp.2018.01.002> (2018).
36. Kushnerenko, E. *et al.* Brain responses and looking behavior during audiovisual speech integration in infants predict auditory speech comprehension in the second year of life. *Front. Psychol.* **4**, <https://doi.org/10.3389/fpsyg.2013.00432> (2013).
37. Kushnerenko, E. *et al.* Brain responses to audiovisual speech mismatch in infants are associated with individual differences in looking behaviour. *Eur. J. Neurosci.* **38**, 3363–3369 <https://doi.org/10.1111/ejn.12317> (2013).
38. Imafuku, M., Kanakogi, Y., Butler, D., & Myowa, M. Demystifying infant vocal imitation: The roles of mouth looking and speaker's gaze. *Dev. Sci.* **22**, <https://doi.org/10.1111/desc.12825> (2019).
39. Polka, L., & Werker, J. F. Developmental Changes in Perception of Nonnative Vowel Contrasts. *J. Exp. Psychol. Hum. Percept. Perform.* **20**, 421–435, <https://doi.org/10.1037/0096-1523.20.2.421> (1994).
40. Rescorla, L., Constants, H., Bialecka-Pikul, M., Stępień-Nycz, M., & Ochał, A. Polish vocabulary development in 2-year-olds: Comparisons with English using the language

development survey. *J. Speech, Lang. Hear. Res.* **60**, 1029–1035, https://doi.org/10.1044/2016_JSLHR-L-15-0385 (2017).

Figures

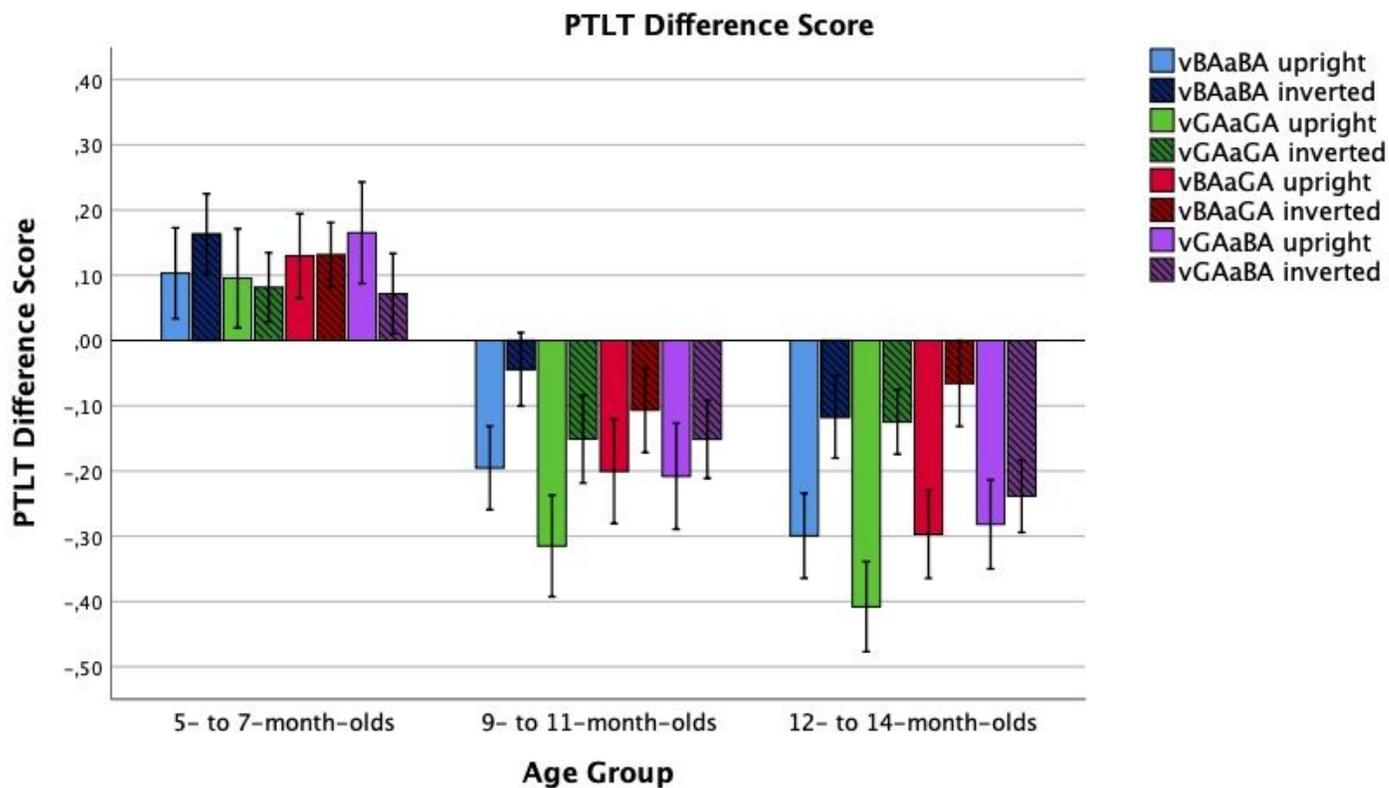


Figure 1

Proportion of Total Looking Time difference scores in each condition and age group. Blue and green bars represent congruent, while red and purple – incongruent AV speech pairs. Error bars represent +/- one standard error of mean.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- [SupplementaryInformationAttentiontospeakingmouthisreducedbyfaceinversionininfantsonlyfrom9monthsofage.pdf](#)