Mask-Wearing Affects Infants' Selective Attention to Familiar and Unfamiliar Audiovisual Speech

Lauren N. Slivka[1], Kenna R.H. Clayton[1], & Greg D. Reynolds[1*]

[1]University of Tennessee, Knoxville, Tennessee, United States

[*]Correspondence:
Greg D. Reynolds
greynolds@utk.edu

infancy, audiovisual speech, mask-wearing, selective attention, eye tracking

ABSTRACT

This study examined the immediate effects of mask-wearing on infant selective visual attention to audiovisual speech in familiar and unfamiliar languages. Infants distribute their selective attention to regions of a speaker's face differentially based on their age and language experience. However, the potential impact wearing a face mask may have on infants' selective attention to audiovisual speech has not been systematically studied. We utilized eye tracking to examine the proportion of infant looking time to the eyes and mouth of a masked or unmasked actress speaking in a familiar or unfamiliar language. Six-month-old and 12-month-old infants (n=42, 55% female, 91% White Non-Hispanic/Latino) were shown videos of an actress speaking in a familiar language (English) with and without a mask on, as well as videos of the same actress speaking in an unfamiliar language (German) with and without a mask. Overall, infants spent more time looking at the unmasked presentations compared to the masked presentations. Regardless of language familiarity or age, infants spent more time looking at the mouth area of an unmasked speaker and they spent more time looking at the eyes of a masked speaker. These findings indicate mask-wearing has immediate effects on the distribution of infant selective attention to different areas of the face of a speaker during audiovisual speech.

1. INTRODUCTION

Faces are ubiquitous in the daily life of infants, accounting for up to 25% of their looking time during waking hours in early infancy (Sudgen et al., 2014). Infants typically experience faces in multimodal contexts during face-to-face interactions involving audiovisual speech. Past research has examined the influence of several factors, including language familiarity, on the distribution of infants' selective attention to audiovisual speech (see review, Bastianello et al., 2022). A relatively consistent trend has been identified across studies with monolingual infants showing that, from around 4 to 8 months of age, infants shift from primarily focusing on the eyes to focusing more on the mouth of a speaker regardless of language familiarity (Lewkowicz & Hansen-Tift, 2012; Pons et al., 2015). This trend is followed by a shift at around 12 months of age toward a more even distribution of looking to the eyes and mouth for a familiar language while a greater focus on the mouth is maintained for an unfamiliar language (Hillairet de Boisferon et al., 2017; Lewkowicz & Hansen-Tift, 2012; Morin-Lessard et al., 2019; Pons et al., 2015, 2019).

The *language expertise hypothesis* (Hillairet de Boisferon et al., 2017; Lewkowicz & Hansen-Tift, 2012; Morin-Lessard et al., 2019; Pons et al., 2015) proposes that the shift back toward the eyes that occurs at around 12 months of age for a familiar language (but not for unfamiliar language) reflects less reliance on the mouth for language processing due to further development of expertise in the familiar language. However, some studies have found that 12-month-olds continue to focus more on the mouth of familiar language speakers (Roth et al., 2022; Tenenbaum et al., 2015; Tsang et al., 2018). Consistent with the proposal that infants focus on the mouth in more demanding language contexts, bilingual infants have been found to shift their focus to the mouth at an earlier age than monolingual infants (Morin-Lessard et al., 2019) and maintain focus on the mouth beyond 12 months of age (Pons et al., 2015).

Infants' patterns of selective attention to audiovisual speech are associated with later language development (Pascalis et al., 2014; Pons et al., 2019; Santapuram et al., 2022). Sensitivity to temporal synchrony in audiovisual speech (but not in non-social stimuli) at 6 months of age is linked with language outcomes at 18, 24, and 36 months of age (Edgar et al., 2023). Greater looking to the eyes at 5.5 months of age correlates with expressive and receptive language in toddlerhood (Lozano et al., 2022). However, greater looking to the mouth relative to the eyes of a speaker between the ages of 6 to 12 months is linked with higher expressive language skills, and greater looking to the eyes at 12 months is linked with higher scores on social and communication subsets of the Bayley Scales of Infant and Toddler Development (Lozano et al., 2022; Morin-Lessard et al., 2019; Pons et al., 2019; Tennebaum et. al., 2019; Tsang et al., 2018; Young et. al., 2009). Overall, patterns of selective attention to audiovisual speech are complex and vary depending on many factors, including: language (Birulés et al., 2019; Sekiyama et al., 2021), monolingualism/bilingualism (Morin-Lessard et al., 2019), prosody (Roth et al., 2022), likelihood of autism (Lozano et al., 2024), sex (Lozano et al., 2022), and term status (Berdasco-Muñoz et al., 2019).

Children born within the COVID-19 cohort (2020-2021) lacked the level of social engagement and learning opportunities available to children born in non-pandemic times (Deoni et al., 2021). Increased use of face masks created a unique problem for infants and children who gather information from faces for social and emotional processing. While mask-wearing is currently not as prevalent in most settings, the potential need for widespread mask-wearing and social distancing during future pandemic outbreaks remains a strong possibility. There is a growing body of research on effects of mask-wearing on face processing in infants and children.

In 2021, Yates and Lewkowicz tested 4-,5-, and 6-year-old children to examine the potential impact of mask-wearing on the developmental timeline of holistic face processing. All three age groups showed evidence of holistic processing, suggesting exposure to segregated visual faces did not disrupt their normal face processing abilities (Yates & Lewkowicz, 2023). However, Ruba and Pollak (2020) found that children 7-11 years of age perform at chance levels at determining the emotion of masked faces. Kammermeier and Paulus (2023) found 14-month-olds are more likely to show an appropriate change in affect when interacting with an unmasked person in comparison to when interacting with a masked person. In adults, mask-wearing leads to an increased use of eye cues to gauge emotions between faces (Barrick et al., 2021). DeBolt and Oakes (2023) explored the impact of mask-wearing on infant memory of faces. Regardless of whether they were familiarized with a masked or unmasked face, 6- and 9-month-olds only demonstrated memory for unmasked faces during testing. This suggests that the effects of mask-wearing on infant face processing are complex and multifaceted. However, the potential effect of mask-wearing on infant attention to audiovisual speech has not been systematically studied.

We designed the current study to examine immediate effects of an adult speaker's mask-wearing on 6- and 12-month-old infants' selective attention to familiar and unfamiliar audiovisual speech. We utilized eye tracking to explore whether infants' proportion of looking time to the eyes and mouth on talking faces varies depending on whether they are masked or unmasked and whether they are speaking a familiar or unfamiliar language. We selected 6 and 12 months as the age groups for this study based on studies showing that infants shift the distribution of their selective attention from primarily focusing on the eyes, or showing equivalent looking to the eyes and mouth at 6 months regardless of language, to a greater focus on the mouth at 12 months for an unfamiliar language but not for a familiar language (Berdasco-Muñoz et al., 2019; Birulés et al., 2019; Hillairet de Boisferon et al., 2017; Lewkowicz & Hansen-Tift, 2012; Lozano et al., 2022; Sekiyama et al., 2021).

Based on the *language expertise hypothesis*, across mask-wearing and language conditions, we expected 6-month-olds to primarily focus their visual attention on the eyes of the speaker relative to the mouth regardless of language familiarity or mask-wearing (H1). We expected 12-month-olds to show differences in proportion of looking time to the eyes and mouth based on language and mask-wearing. We predicted 12-month-olds in the unmasked condition would focus more on the eyes in the familiar language condition in comparison to the unfamiliar language condition (H2) and more on the mouth for the unfamiliar language in comparison to familiar language (H3). In the masked condition, we predicted, regardless of language familiarity, infants would primarily focus their visual attention on the eyes of the speaker relative to the mouth (H4).

2. METHODS

2.1 Participants

Fifty-two infants were recruited from the Knoxville, TN area between 2022-2023. Ten infants were excluded from the final data set due to fussiness (*n*=4), failure to calibrate (*n*=2), and failure to have data on a minimum number of trials (*n*=4). The final sample (*N*=42) included a group of 6-month-old infants (*n*=22, *Mage*=179.33 days, *SD*=7.09 days, 12 females) and a group of 12-month-old infants (*n*=20, *Mage*=358.41 days, *SD* = 8.27 days, 11 females). The racial distribution of the final sample was 38 White, 1 Black, and 3 mixed-race infants. All infants were born full-term (at least 37 weeks of gestation) with no known visual or hearing difficulties. Participants were monolingual, English-learning infants. Participants' caregivers reported exposure to German in their household or elsewhere was little to none (<10%). All infants were tested within plus or minus two weeks before or after their 6- or 12-month birthdate. The guardians received a certification of participation, an infant sized t-shirt, and a small cash payment for participation. Only infants that provided useable data for a minimum of one presentation of each condition (see Audiovisual Stimuli below) were included in the analysis.

2.2 Apparatus

Participants sat in their guardian's lap, with the position of their eyes centered vertically and horizontally on the display monitor. The infants were positioned approximately 55 cm away from the display monitor (17" Dell UltraSharp 1704FPV color LCD monitor). Speakers were placed behind the monitor, hidden from view of the infants, to play the accompanying soundtrack to the video clips. Testing took place in a sound attenuated room with black curtains surrounding the testing area. A remote eye-tracker (SR Research Ltd., Eye-Link 1000 plus) was placed under the LCD monitor facing the infant. Infants' gaze was recorded at a sampling rate of 500 Hz.

2.3 Audiovisual Stimuli

The stimuli were 30-second continuous audiovisual video clips of the same actress speaking in either German or English with the audio presented at 60 dB at the position where the infant was seated during testing. The actress had her hair pulled back, wore a plain black t-shirt, and was filmed against a plain background. All four videos included the same actress (White, unbalanced bilingual with primary language being English, aged 28) in the same filming location, reciting the same script (See Figure 1 for examples of the masked and unmasked actress). For the familiar language condition, the actress spoke in English. For the unfamiliar language condition, the actress recited the same script as the familiar condition, but in German. The English and German script were both spoken in infant directed speech (IDS). For the masked conditions, the actress recited the script wearing a blue surgical mask. Infants viewed two blocks for each of the four conditions (i.e., English unmasked, English masked, German unmasked, and German masked).

2.4 Procedure

Informed consent was completed with the participants' guardian prior to the start of testing. The infant was placed in the guardian's lap facing the monitor with their eyes level with the center of the monitor. The guardian was asked to refrain from talking to or moving the infant during the study. A bullseye tracking sticker was placed on the infant's forehead to aid the eye tracker in pupil localization. Curtains were drawn around the infant with the lights dimmed. A 5-point calibration routine was then performed to ensure accurate eye tracking. A random selection of a donut or geometric shape with sound moved from the bottom-right, bottom-left, top-left, top-right, and center of the screen, subtending a visual angle of 2.32°. These calibration points were repeated until the infant attended to all five of these points and a validation each point ensured adequate tracking of the infants' pupils. Successful calibration was determined by the infants attending all five points within one visual degree of the calibration point. After calibration, an animated cartoon was played continuously in the center of the screen as an attention-getter until the infant was centrally fixated for the start of the trial. Test stimuli were not presented until the experimenter pressed a button, after judging the infant to be centrally fixated. Stimuli consisted of eight trials, each composed of a 30-second video clip, with an attention-getter shown between each trial. Order of presentation (masked or unmasked and German or English first) was randomized between participants to counterbalance for potential order effects (see Figure 1 for a diagram of a trial block).

2.5 *Measurement*

The stimuli included areas of interest (AOIs): one around the bottom half of the speaker's face (defined as mouth AOI, 44095 pixels), and one around the top half of the speaker's face (defined as eyes AOI, 44096 pixels) (see Panel B on Figure 1). When viewed from 55cm away, test stimuli subtended a visual angle of approximately 28.49° x 32.10°. These AOIs were defined using DataViewer (SR Research, Ltd) to equally split the speaker's face from the center up to include the eyes, and the center down to include the mouth. Looks that occurred outside these AOIs were discarded from analysis during data processing. We analyzed total dwell time to the face as an index of level of attention. Total dwell time was defined as the amount of time in milliseconds (ms) the infant spent looking anywhere within the face region. We also analyzed the proportion of total looking time (PTLT) to each of the AOIs. PTLT was defined as the dwell time to a specific AOI (mouth or eyes) divided by the overall dwell time to the entire face. To derive our primary dependent variable, PTLT difference scores, we subtracted PTLT to the mouth from PTLT to the eyes for each condition. This approach allowed us to quantify a relative preference for gaze allocation with a positive PTLT difference indicating greater proportion of looking time to the eyes and a negative score indicating greater proportion of looking time to the mouth. Any participant who failed to contribute looking data (i.e., at least two fixations with each fixation exceeding a minimum duration of 66ms) for at least one presentation of each condition (English, English Masked, German, German Masked) was excluded from analysis.

3. RESULTS

This study utilized a 2x2x2 mixed design with age (6- and 12-months) as a between-subjects factor and language familiarity (familiar or unfamiliar), and mask-wearing condition (masked or unmasked) as within-subjects factors. Repeated measures ANOVA (Analysis of Variance) were used to identify significant experimental effects. Using Bonferroni's correction, post-hoc analyses using one-way ANOVAs and paired sample t-tests were conducted with alpha levels set to 0.05. To examine whether infants demonstrated a preference for the eyes or the mouth based on experimental factors, single-sample t-tests were carried out across conditions to determine if greater proportion of looking time to an AOI exceeded a chance value (0.0).

Using PTLT difference to examine differences in infants' PTLT to each AOI, a repeated measures ANOVA was conducted with Age as a between-subjects factor (6 and 12 months) and Language Familiarity (German and English) and Mask-Wearing (Masked and Unmasked) as within-subject factors. Means and individual participant scores on PTLT difference for all conditions are shown in Figure 2 and means with standard deviations for all conditions are shown in Table 1. The ANOVA table for the full factorial ANOVA is presented in Table 2 for the interested reader (see Supplemental Material).

There was a significant main effect for Mask-Wearing, $F(1, 40) = 57.478$, $p < 0.001$, $\eta_p^2 = 0.590$ (see Figures 2 and 3). As shown by a one-sample t-test comparing PTLT difference to chance (0.0), infants looked more to the mouth than the eyes of the unmasked speaker ($M = -0.201$, $SD = .429$; $t(41) = -3.033$, $p < .01$, $d = -0.468$). In contrast, in the masked condition, infants looked more to the eyes than the mouth of the speaker ($M = 0.247$, $SD = .481$; $t(41) = 3.336$, $p < .001$, $d = 0.515$) regardless of age or language. There were no other significant main effects or interactions between factors (all ps > .10, all $\eta_p^2 < .10$).

To further examine whether the infants' PTLT difference represented a significant preference for an AOI (e.g., Morin-Lessard et al., 2019), one-sample t-tests were carried out for each condition and age group comparing PTLT difference to chance (0.0). At 6 months of age, PTLT difference in the unmasked conditions did not differ from chance; however, in the masked conditions, PTLT difference exceeded chance with greater proportion of looking to the eyes for both familiar language, $t(21) = 2.411$, $p < .05$, $d = 0.592$, and unfamiliar language, $t(21)= 3.987$, $p < .001$, $d = 0.850$. At 12 months of age, PTLT difference for the unmasked, unfamiliar language condition was the only condition that exceeded chance levels with greater proportion of looking to the mouth, $t(19) = -2.638$, $p < .01$, $d = -0.590$.

Total dwell time, defined as the total amount of time in ms spent looking at the entire face, was analyzed to determine if there were any differences in level of attention between the familiar and unfamiliar language conditions as well as the masked and unmasked conditions. No difference in total dwell time was found between familiar language ($M = 10993.17$ ms, $SD = 7003.00$) and unfamiliar language trials ($M = 10381.83$ ms, $SD = 6416.62$), $t(41) = 1341$, $p = .094$. However, infants had higher total dwell times on unmasked speaker trials ($M = 11854.29$ ms, $SD = 7397.63$) compared to masked speaker trials ($M = 9520.71$ ms, $SD = 6245.12$), $t(41)=-3.811$, $p < .001$, $d = 0.588$.

4. DISCUSSION

This study explored whether infants' selective attention to facial features differs between audiovisual speech from a familiar or unfamiliar language spoken by a masked or unmasked talking face. There was a main effect of mask-wearing. Infants looked proportionately more to the mouth of the speaker in the unmasked condition and more to the eyes of the speaker in the masked condition. We predicted (H1) 6-month-old infants would look proportionately more to the eyes of the speaker compared to the mouth, regardless of language familiarity or mask-wearing condition. In contrast, we found that 6-month-olds only looked more to the eyes of a masked speaker. In the unmasked condition, 6-month-olds' proportion of looking time did not differ significantly between the eyes and the mouth. This finding is somewhat unexpected for the unmasked condition given that past work has found that 5- to 6-month-olds look longer to the eyes of familiar and unfamiliar language speakers (e.g., Berdasco-Muñoz et al., 2019; Hillairet de Boisferon et al., 2017; Lozano et al., 2022; Sekiyama et al., 2021). However, some studies have also found that 5- to 6-month-olds show equivalent looking to the eyes and mouth of a speaker (Lewkowicz & Hansen-Tift, 2012; Morin-Lessard et al., 2019) indicating that proportion of looking time to the eyes and mouth is likely influenced by multiple factors at this age. To equate the size of AOIs for our mask-wearing conditions, we defined the AOIs for the eyes and mouth as the upper and lower half of the face, respectively. Most studies have used AOIs covering more narrow regions of the face specific to each of these facial features. Thus, the larger AOIs used in the current study may have contributed to the equivalent looking to the eyes and mouth AOIs for 6-month-olds.

Based on prior work and the *language expertise hypothesis* (Hillairet De Boisferon et al., 2017; Lewkowicz & Hansen-Tift, 2012; Pons et al. 2015), we predicted 12-month-olds in the unmasked condition would focus more on the eyes in the familiar language condition in comparison to the unfamiliar language condition (H2) and more on the mouth for the unfamiliar language in comparison to familiar language (H3). Comparisons across conditions did not support these predictions. There were no differences in PTLT difference between the familiar and unfamiliar language conditions. However, when comparing PTLT difference with chance to examine preference for the mouth or eyes within each condition, the only PTLT difference that was significantly greater than chance at 12 months of age was the greater looking to the mouth than the eyes of the unmasked, unfamiliar language speaker. This finding, combined with the lack of language familiarity effects on PTLT difference at 6 months of age, is consistent with the *language expertise hypothesis* (Hillairet De Boisferon et al., 2017; Lewkowicz & Hansen-Tift, 2012; Pons et al. 2015), and indicates 12-month-olds rely heavily on the mouth of an unmasked speaker when being spoken to in an unfamiliar language.

We used German as the unfamiliar language and English as the familiar language. Given that English is a Germanic language, English is structurally and prosodically close to German. It is possible that the closeness between the two languages used for the current study accounts for the lack of significant differences *between* language familiarity conditions on proportion of looking time to the eyes and mouth. Prior research has shown that 15-month-old close-language learning bilingual infants attend more to the mouth of speakers than 15-month-old distant-language learning bilingual infants (Birulés et. al., 2019), indicating that language-closeness does influence the distribution of infant selective attention to the mouth of a speaker. Further, using a balanced bilingual for future studies would be ideal. The actress used for this study was not a native German speaker which may have resulted in greater similarity in speech across the familiar and unfamiliar languages.

Our final prediction (H4) was that 12-month-olds would look more to the eyes than the mouth of a masked speaker regardless of language familiarity. The main effect of mask-wearing supported this prediction; regardless of language familiarity, infants looked longer to the mouth than the eyes of the speaker when unmasked (similar to, Roth et al., 2022; Tenenbaum et al., 2015; Tsang et al., 2018) and longer to the eyes than the mouth of the speaker when masked. Although this study was focused on the immediate effects of mask-wearing on infant selective attention to audiovisual speech, the current findings could have important implications for infants raised during an active pandemic with widespread mask-wearing. The 6-to-12-month age range is a formative period in language development. Research links infants' increased attention to the mouth of a communicative partner with higher expressive communication, higher vocal complexity, and higher scores on the BSID-II in toddlerhood (Santapuram et al., 2022; Pons et al., 2019; Tsang et al., 2018). Attentional preference to the mouth of a communicative partner between the ages of 6- and 12-months of age is positively associated with expressive language skills, and an attentional preference to the eyes at 12 months of age is positively associated with social and communication subsets of the BSID-III (Pons et al., 2019; Tsang et al., 2018). Sensitivity to temporal synchrony in audiovisual speech at 6 months of age is linked with language outcomes at 18-, 24-, and 36-months of age (Edgar et al., 2023). Furthermore, research indicates infants only show evidence of facial recognition when the faces are unmasked during testing (DeBolt & Oakes, 2023). Further research is needed to examine the potential impact of the effects of wearing masks on infants' selective attention to audiovisual speech on early language learning and social development.

In addition to looking proportionately less to the mouth of a masked speaker in comparison to an unmasked speaker, our analysis of total dwell time to the entire face revealed that infants paid less attention overall to a masked speaker in comparison to an unmasked speaker. Across 6 to 12 months of age, infants pay greater attention to more complex and dynamic stimuli than to more basic and static stimuli (e.g., Courage et al., 2006; Reynolds et al., 2013). Thus, the lower overall attention in the masked conditions was likely due to the reduced complexity (i.e., fewer internal elements) and reduced motion associated with occlusion of the mouth and nose.

There were some notable limitations to the current study. Recruitment constraints resulted in a relatively homogenous (i.e., primarily White) racial/ethnic distribution for the final sample. Future research is needed to examine these research questions across a larger and more diverse sample of participants. We utilized a cross-sectional design to examine immediate effects of mask-wearing on infant attention to audiovisual speech. Under the appropriate context, a longitudinal design could highlight developmental effects of mask-wearing on selective attention to audiovisual speech as well as examining other outcome measures. Additionally, we did not collect mask exposure information. The infants tested in this study were born after lifting of mask mandates tied to COVID-19 and mask-wearing had become relatively uncommon in the local community. The current findings reflect immediate effects of mask-wearing on selective attention to audiovisual speech for infants who likely had relatively little prior exposure to masked faces. Thus, any implications of these findings for potential developmental effects of long-term exposure to masked faces in infancy are purely speculative.

4.2 Conclusions

The current study is an initial step in understanding how infants shift their selective attention to facial features in real-time when exposed to either masked or unmasked speakers of a familiar or unfamiliar language. Infants looked proportionately more at the eyes and less at the mouth of a masked speaker, which is the opposite of the pattern of selective attention shown in response to an unmasked speaker. Infants were also less attentive overall to audiovisual speech when the speaker was wearing a mask. Past research highlights the importance of attending to the mouth during language development in mid- to late-infancy as well as the importance of distributing selective attention to the eyes for later social development (e.g., Pons et al., 2019; Santapuram et al., 2022; Schwarzer et al., 2007; Tsang et al., 2018). Developmental implications of these effects under conditions of widespread face mask usage during a pandemic remain speculative but may warrant consideration of the use of transparent face masks that do not occlude the view of the mouth in early childcare and educational settings.

Conflict of Interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Author Contributions

LS was primarily responsible for data collection and contributed to study design, analysis, and writing of the manuscript. KC contributed to data collection and analysis. GR provided oversight for the execution of the project and contributed to study design, analysis, and writing of the manuscript.

Funding

This material is based upon work supported by the National Science Foundation under Grant No. BCS- 2217766 awarded to GR. Support for this work was also provided by a Student/Faculty Research Award to LS and GR from the Graduate School of the University of Tennessee. Funding for open access to this research was provided by University of Tennessee's Open Publishing Support Fund.

Acknowledgments

The authors would like to express their sincere gratitude to the infant participants and their guardians for their invaluable contributions to this study.

References

Barrick, E. M., Thornton, M. A., & Tamir, D. I. (2021). Mask exposure during COVID-19 changes emotional face processing. *PloS one*, *16*(10), e0258470. https://doi.org/10.1371/journal.pone.0258470

Bastianello, T., Keren-Portnoy, T., Majorano, M., & Vihman, M. (2022). Infant looking preferences towards dynamic faces: A systematic review. *Infant Behavior and Development*, *67*, 101709. https://doi.org/10.1016/j.infbeh.2022.101709

Berdasco-Muñoz, E., Nazzi, T., & Yeung, H. H. (2019). Visual scanning of a talking face in preterm and full-term infants. *Developmental Psychology*, *55*(7), 1353. https://doi.org/10.1037/dev0000737

Birulés, J., Bosch, L., Brieke, R., Pons, F., & Lewkowicz, D. J. (2019). Inside bilingualism: Language background modulates selective attention to a talker's mouth. *Developmental Science*, *22*(3), e12755. https://doi.org/10.1111/desc.12755

Courage, M. L., Reynolds, G. D., & Richards, J. E. (2006). Infants' Attention to Patterned Stimuli: Developmental Change From 3 to 12 Months of Age. Child Development, 77(3), 680–695. https://doi.org/10.1111/j.1467-8624.2006.00897.x

DeBolt, M. C., & Oakes, L. M. (2023). The impact of face masks on infants' learning of faces: An eye tracking study. *Infancy*, *28*(1), 71–91. https://doi.org/10.1111/infa.12516

Deoni, S. C., Beauchemin, J., Volpe, A., D'Sa, V., & the RESONANCE Consortium. (2021). *Impact of the COVID-19 Pandemic on Early Child Cognitive Development: Initial Findings in a Longitudinal Observational Study of Child Health* [Preprint]. Pediatrics. https://doi.org/10.1101/2021.08.10.21261846

Edgar, E. V., Todd, J. T., & Bahrick, L. E. (2023). Intersensory processing of faces and voices at 6 months predicts language outcomes at 18, 24, and 36 months of age. *Infancy*, *28*(3), 569–596. https://doi.org/10.1111/infa.12533Flom, R., & Bahrick, L. E. (2007). The development of infant discrimination of affect in multimodal and unimodal stimulation: The role of intersensory redundancy. *Developmental Psychology*, *43*(1), 238–252. https://doi.org/10.1037/0012-1649.43.1.238

Kammermeier, M., & Paulus, M. (2023). Infants’ responses to masked and unmasked smiling faces: A longitudinal investigation of social interaction during Covid-19. *Infant Behavior and Development*, *73*, 101873. https://doi.org/10.1016/j.infbeh.2023.101873

Lewkowicz, D. J., & Hansen-Tift, A. M. (2012). Infants deploy selective attention to the mouth of a talking face when learning speech. *Proceedings of the National Academy of Sciences*, *109*(5), 1431–1436. https://doi.org/10.1073/pnas.1114783109

Lozano, I., Belinchón, M., & Campos, R. (2024). Sensitivity to temporal synchrony and selective attention in audiovisual speech in infants at elevated likelihood for autism: A preliminary longitudinal study. Infant Behavior and Development, 76, 101973. https://doi.org/10.1016/j.infbeh.2024.101973

Lozano, I., López Pérez, D., Laudańska, Z., Malinowska-Korczak, A., Szmytke, M., Radkowska, A., & Tomalski, P. (2022). Changes in selective attention to articulating mouth across infancy: Sex differences and associations with language outcomes. Infancy, 27(6), 1132–1153. https://doi.org/10.1111/infa.12496

Maurer, D., Grand, R. L., & Mondloch, C. J. (2002). The many faces of configural processing. *Trends in Cognitive Sciences*, *6*(6), 255–260. https://doi.org/10.1016/S1364-6613(02)01903-4

Morin-Lessard, E., Poulin-Dubois, D., Segalowitz, N., & Byers-Heinlein, K. (2019). Selective Attention to the Mouth of Talking Faces in Monolinguals and Bilinguals Aged 5 Months to 5 Years. Developmental Psychology, 55(8), 1640–1655. https://doi.org/10.1037/dev0000750

Pascalis, O., Loevenbruck, H., Quinn, P. C., Kandel, S., Tanaka, J. W., & Lee, K. (2014). On the Links Among Face Processing, Language Processing, and Narrowing During Development. *Child Development Perspectives*, *8*(2), 65–70. https://doi.org/10.1111/cdep.12064

Pons, F., Bosch, L., & Lewkowicz, D. (2015). Bilingualism modulates infants' _selective attention to the mouth of a talking face. *Psych. Sci., 26*, 490–498. https://doi. org/10.1177/0956797614568320

Pons, F., Bosch, L., & Lewkowicz, D. J. (2019). Twelve-month-old infants' attention to the eyes of a talking face is associated with communication and social skills. *Infant Behavior and Development*, *54*, 80–84. https://doi.org/10.1016/j.infbeh.2018.12.003

Pons, F., & Lewkowicz, D. J. (2014). Infant perception of audio-visual speech synchrony in familiar and unfamiliar fluent speech. *Acta Psychologica*, *149*, 142–147. https://doi.org/10.1016/j.actpsy.2013.12.013

Reynolds, G. D., Zhang, D., & Guy, M. W. (2013). Infant attention to dynamic audiovisual Stimuli: Look duration from 3 to 9 months of age. Infancy, 18(4), 554–577. https://doi.org/10.1111/j.1532-7078.2012.00134.x

Roth, K. C., Clayton, K. R. H., & Reynolds, G. D. (2022). Infant selective attention to native and non-native audiovisual speech. *Scientific Reports*, *12*(1), 15781. https://doi.org/10.1038/s41598-022-19704-5

Ruba, A. L., & Pollak, S. D. (2020). Children's emotion inferences from masked faces: Implications for social interactions during COVID-19. *PLOS ONE*, *15*(12), e0243708. https://doi.org/10.1371/journal.pone.0243708

Santapuram, P., Feldman, J. I., Bowman, S. M., Raj, S., Suzman, E., Crowley, S., Kim, S. Y., Keceli-Kaysili, B., Bottema-Beutel, K., Lewkowicz, D. J., Wallace, M. T., & Woynaroski, T. G. (2022). Mechanisms by Which Early Eye Gaze to the Mouth During Multisensory Speech Influences Expressive Communication Development in Infant Siblings of Children with and Without Autism. *Mind, Brain, and Education*, *16*(1), 62–74. https://doi.org/10.1111/mbe.12310

Schwarzer, G., Zauner, N., & Jovanovic, B. (2007). Evidence of a shift from featural to configural face processing in infancy. *Developmental Science*, *10*(4), 452–463. https://doi.org/10.1111/j.1467-7687.2007.00599.x

Sekiyama, K., Hisanaga, S., & Mugitani, R. (2021). Selective attention to the mouth of a talker in Japanese-learning infants and toddlers: Its relationship with vocabulary and compensation for noise. *Cortex*, *140*, 145–156. https://doi.org/10.1016/j.cortex.2021.03.023

Tenenbaum, E. J., Sobel, D. M., Sheinkopf, S. J., Malle, B. F., & Morgan, J. L. (2015). Attention to the mouth and gaze following in infancy predict language development. *Journal of Child Language*, *42*(6), 1173–1190. https://doi.org/10.1017/S0305000914000725

Tsang, T., Atagi, N., & Johnson, S. P. (2018). Selective attention to the mouth is associated with expressive language skills in monolingual and bilingual infants. *Journal of Experimental Child Psychology*, *169*, 93–109. https://doi.org/10.1016/j.jecp.2018.01.002

Yates, T. S., & Lewkowicz, D. J. (2023). Robust holistic face processing in early childhood during the COVID-19 pandemic. *Journal of Experimental Child Psychology*, *232*, 105676. https://doi.org/10.1016/j.jecp.2023.105676

Young, G. S., Merin, N., Rogers, S. J., & Ozonoff, S. (2009). Gaze behavior and affect at 6 months: Predicting clinical outcomes and language development in typically developing infants and infants at risk for autism. *Developmental Science*, *12*(5), 798–814. https://doi.org/10.1111/j.1467-7687.2009.00833.x

Figure Legends

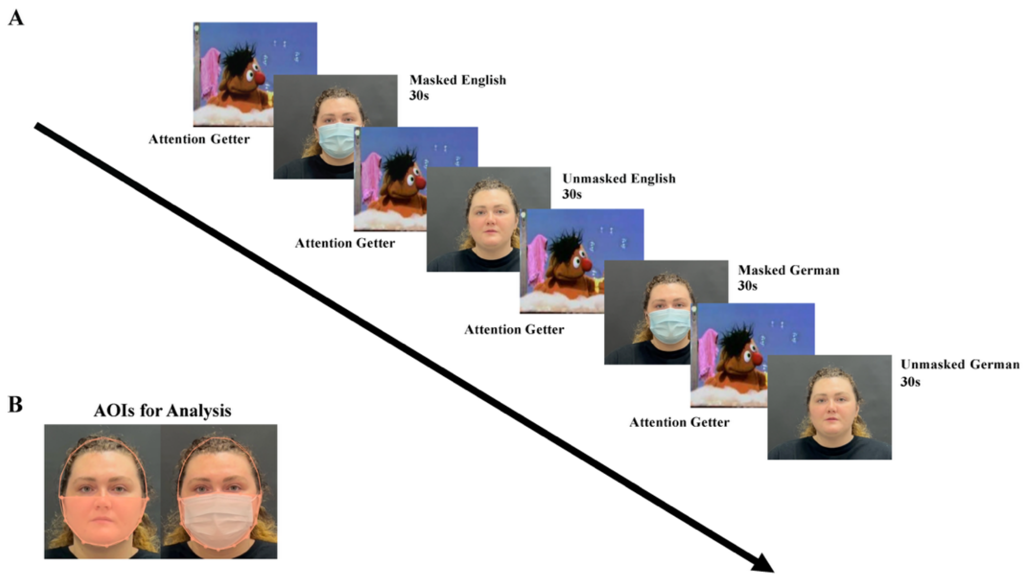
Figure 1: Panel A: presentation sequence of a trial block beginning with an attention getter followed by each of the 4 video presentations for each condition. Each mask-wearing condition was presented for one of the languages, followed by each mask-wearing condition for the other language. Order of language and mask-wearing condition was counterbalanced across participants. Each participant viewed two blocks of the same sequence of presentations. Panel B shows displays screengrabs from presentations of each mask-wearing condition. Orange outlines represent location of facial AOIs used for the analysis of PTLT. Shaded area denotes "mouth" AOI, and the unshaded area represents the "eyes" AOI.

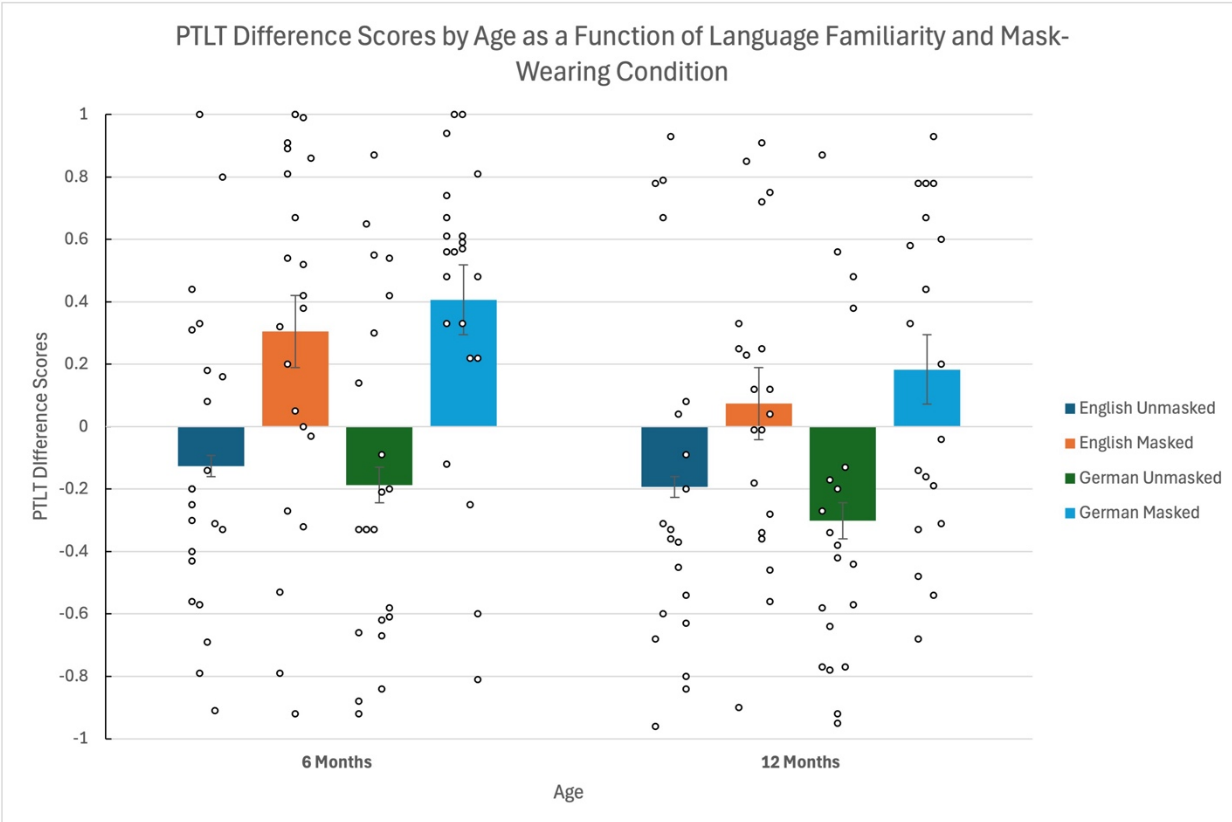
Figure 2: PTLT (proportion of total looking time) Difference Scores by Age as a Function of Language and Mask-Wearing Condition. The y-axis represents AOI difference scores with positive scores indicating greater proportion of looking time to the eyes and negative scores indicating greater proportion of looking time to the mouth. Age is represented on the x-axis. Language and masking-wearing conditions are represented as separate color bars.

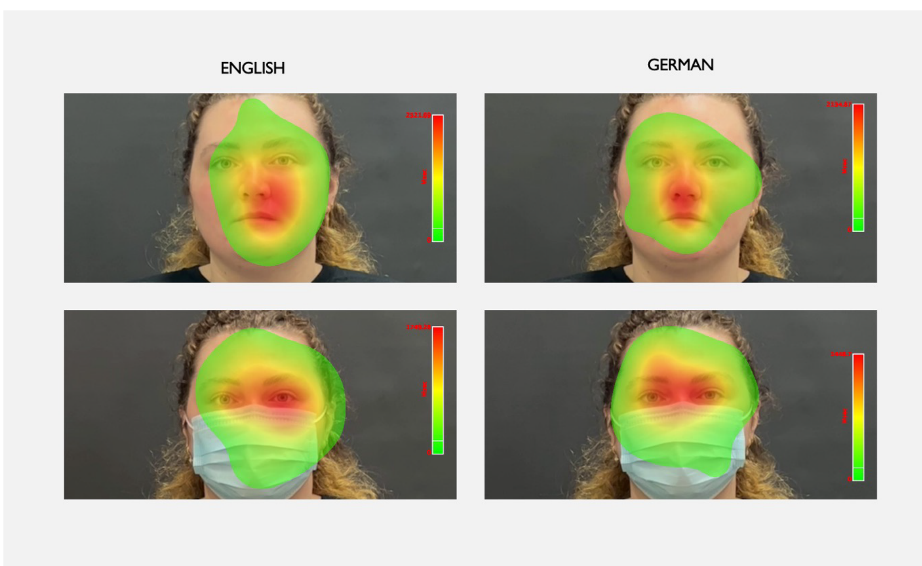
Figure 3: Heatmaps displaying aggregated raw gaze across stimulus types. The heatmaps are averaged across age groups. English presentations (masked & unmasked) are represented within the left panel while German presentations (masked & unmasked) are represented within the right panel.

Table 1. Means and Standard Deviations of PTLT Difference Scores by Function of Age, Language, and Mask-Wearing Condition

	Age	Mean	Std. Deviation
English Unmasked	6	-.1257	.49419
	12	-.1949	.57637
	Average	-.1587	.52940
English Masked	6	.3041	.59170
	12	.0737	.48895
	Average	.1944	.55107
German Unmasked	6	-.1884	.53913
	12	-.3031	.51382
	Average	-.2430	.52401
German Masked	6	.4070	.47878
	12	.1831	.51236
	Average	.3004	.50188


A
Masked English
30s
Attention Getter
Unmasked English
30s
Attention Getter
Masked German
30s
Attention Getter
Unmasked German
30s
Attention Getter
B
AOIs for Analysis


PTLT Difference Scores by Age as a Function of Language Familiarity and Mask-Wearing Condition
PTLT Difference Scores
1
0.8
0.6
0.4
0.2
0
-0.2
-0.4
-0.6
-0.8
-1
6 Months
12 Months
Age
English Unmasked
English Masked
German Unmasked
German Masked


ENGLISH
GERMAN