Brain & Language 244 (2023) 105304

Contents lists available at ScienceDirect

Brain and Language

journal homepage: www.elsevier.com/locate/b&l

Watching talking faces: The development of cortical representation of visual syllables in infancy

Aleksandra A.W. Dopierała [a,b,*], David López Pérez [c], Evelyne Mercure [d], Agnieszka Pluta [a,g], Anna Malinowska-Korczak [c], Samuel Evans [e,f], Tomasz Wolak [g], Przemysław Tomalski [a,c,*]

[a] Faculty of Psychology, University of Warsaw, Warsaw, Poland
[b] Department of Psychology, University of British Columbia, Vancouver, Canada
[c] Institute of Psychology, Polish Academy of Sciences, Warsaw, Poland
[d] Goldsmiths, University of London, London, UK
[e] University of Westminister, London, UK
[f] Kings College London, London, UK
[g] Institute of Physiology and Pathology of Hearing, Bioimaging Research Center, World Hearing Centre, Warsaw, Poland

ARTICLE INFO

Keywords:
fNIRS
Visual speech
Infant
Speech processing
Dynamic face processing

ABSTRACT

From birth, we perceive speech by hearing and seeing people talk. In adults cortical representations of visual speech are processed in the putative temporal visual speech area (TVSA), but it remains unknown how these representations develop. We measured infants' cortical responses to silent visual syllables and non-communicative mouth movements using functional Near-Infrared Spectroscopy. Our results indicate that cortical specialisation for visual speech may emerge during infancy. The putative TVSA was active to both visual syllables and gurning around 5 months of age, and more active to gurning than to visual syllables around 10 months of age. Multivariate pattern analysis classification of distinct cortical responses to visual speech and gurning was successful at 10, but not at 5 months of age. These findings imply that cortical representations of visual speech change between 5 and 10 months of age, showing that the putative TVSA is initially broadly tuned and becomes selective with age.

1. Introduction

From the moment they are born, most infants not only hear but also see people talk. Within the first months of life, they can match visual articulations to appropriate speech sounds (Kuhl & Meltzoff, 1982; MacKain et al., 1983; Patterson & Werker, 1999, 2003) and discriminate languages just by viewing silent visual speech (Kubicek et al., 2014; Weikum et al., 2007). Within the second half of the first year of life speech processing changes: Infants gradually become less skilled in discriminating rarely experienced non-native auditory speech sound contrasts (Werker & Tees, 1984). A similar decline in performance is observed for visual speech: Four- and 6-month-olds but not 8-month-olds can discriminate native and non-native languages visually (Weikum et al., 2007). Furthermore, between 5 and 10 months of age infants' visual attention to talking faces changes, likely reflecting increasing knowledge about visemes (Tomalski et al. 2013), the easily distinguishable mouth movements that can be lipread. These behavioural findings show that infants process visual speech but offer limited answers to questions of how visual speech is represented at a cortical level. To our knowledge, no study to date has measured the development of differential responses to visual speech compared to other mouth movements within the first year of life.

1.1. The fronto-temporal network for visual speech

To make predictions of how cortical representations develop in infancy, we first turn to adult studies. In adults visual speech engages a fronto-temporal network: inferior frontal and superior temporal cortices (Bernstein & Liebenthal, 2014). The network is active to both audiovisual and unisensory (auditory and visual) speech (Dick et al., 2010;

* Corresponding authors at: 2136 West Mall, Vancouver, BC V6T 1Z4, Canada (A.A.W. Dopierała). Jaracza 1, 00-378 Warsaw, Poland (P. Tomalski).
E-mail addresses: aleksandra.dopierala@psych.uw.edu.pl, ptomalski@psych.pan.pl (A.A.W. Dopierała), d.lopez@psych.pan.pl (D. López Pérez), E.Mercure@gold.ac.uk (E. Mercure), apluta@psych.uw.edu.pl (A. Pluta), a.malinowska-korczak@psych.pan.pl (A. Malinowska-Korczak), S.Evans1@westminster.ac.uk (S. Evans), t.wolak@ifps.org.pl (T. Wolak), tomalski@mac.com (P. Tomalski).

https://doi.org/10.1016/j.bandl.2023.105304
Received 22 July 2022; Received in revised form 13 July 2023; Accepted 17 July 2023
0093-934X/© 2023 Elsevier Inc. All rights reserved.

Matchin et al., 2014; Ojanen, 2005; Olson et al., 2002; Sekiyama et al., 2003; Venezia et al., 2017). When compared with gurning - i.e., mouth movements that carry no linguistic information - the fronto-temporal network shows higher activation to visual speech (Campbell et al., 2001; Hall et al., 2005; Okada & Hickok, 2009; Rorden et al., 2008). As both visual speech and gurning involve dynamic and configural face processing, the difference in activation to visual speech and gurning can be attributed to the processing of visemic information. The left hemisphere typically shows greater speech-specific activations (visual speech vs gurning) compared to the right hemisphere (e.g., Sekiyama et al., 2003; Venezia et al., 2017). The posterior STS (pSTS) shows an anterior-to-posterior gradient specialisation for facial motion processing. Gurning preferentially activates anterior parts of the pSTS, while visual speech - posterior parts of the pSTS (for a review see Venezia et al., 2017). A functionally specialised cortical region posterior and inferior to the left superior temporal sulcus (STS) - the temporal visual speech area (TVSA, Bernstein et al., 2011; Bernstein & Liebenthal, 2014) - processes cortical representations of visemes. Greater TVSA activation to visual speech than gurning reflects specialisation for viseme processing and is interpreted as evidence for modality-specific cortical representations of visual speech (Bernstein et al., 2011; Bernstein & Liebenthal, 2014). In summary, adults have modality-specific cortical representations of visemes which are distinct from representations of other mouth movements, and processed in the putative TVSA.

While no studies have specifically investigated the development of cortical responses to visual speech, existing functional near-infrared (fNIRS) studies on cortical responses to speech and facial motion in infancy suggest that the fronto-temporal network is involved. Around 5 months of age frontal responses to speech differed depending on modality (visual, auditory, audiovisual) (Altvater-Mackensen & Grossmann, 2018). Additionally, by 5 months of age, infants showed differential fronto-temporal responses to dynamic faces depending on the type of facial motion, such as eye gaze shift, eye-brow raise or mouth movement (Grossmann et al., 2008; Lloyd-Fox et al., 2011). Between 5 and 8 months of age, cortical responses to mouth movements were different depending on familiarity of the movement: Bilateral superior temporal (ST) regions were active to yawning but not to other, unfamiliar and non-communicative mouth movements (Tsurumi et al., 2019). Altogether, these studies demonstrate that the fronto-temporal regions are functionally active in infancy during the perception of audiovisual speech and facial motion. They show that infants have distinct fronto-temporal representations of auditory, visual, and audiovisual vowels, eyes and mouth movements, as well as familiar and unfamiliar mouth movements. However, the specificity of cortical representations of visual speech or how they develop with age remained unclear. No study to date has specifically addressed the sensitivity of the superior temporal cortex and the putative TVSA to visual speech, the selectivity of the fronto-temporal cortices to visual speech compared to gurning, or tested how these change with age.

1.2. This study

To shed more light on the development of cortical representations of visual speech, we present the first cross-sectional investigation of cortical responses to visual speech and gurning in infancy. We used functional Near-Infrared Spectroscopy (fNIRS) to measure the fronto-temporal responses to non-social dynamic stimuli (baseline), visual speech (syllables), and gurning. We included two dynamic visual stimuli (non-social, gurning) to draw conclusions about the specificity of infants' cortical representations of visual speech, i.e., identify cortical regions sensitive to social compared to non-social dynamic stimuli (visual speech versus baseline, gurning versus baseline) and regions differentially active to visual speech and gurning. Our pre-registered hypotheses (Dopierała et al., 2019) predicted that relative to baseline, the processing of visual speech engages the fronto-temporal network - including the inferior frontal and superior temporal cortices and the likely region of the putative TVSA - in both 5- and 10-month-olds. We proposed that the cortical organisation supporting visual speech changes between 5 and 10 months of age, a time when infants' knowledge about visemes likely increases: Infants become increasingly proficient at processing both speech and faces (e.g., Werker & Tees, 1984) and their visual attention to audiovisual speech changes (Lewkowicz & Hansen-Tift, 2012; Tomalski et al., 2013). Around 5 months of age, the putative TVSA likely specialises in the processing of any type of mouth movements and responses are similar for visual syllables and gurning. By 10 months of age, the region specialises to specifically process visible articulatory mouth movements, showing distinct responses to presented mouth movements. In infants, speech processing also shows increasingly left-lateralised responses within the first year of life (Minagawa-Kawai et al., 2007). We predicted that between 5 and 10 months of age, the left hemisphere will become dominant for visual speech. Specifically, left inferior frontal and superior temporal responses to visual speech will be greater in older than younger infants. Similarly, the sensitivity of the putative TVSA to visual speech will also increase with age: The differential response to visual speech versus gurning will be more pronounced in the older than younger age group.

To gain a fuller understanding of the development of cortical specialisation for visual speech, we run both univariate channel-by-channel analyses and Multivariate Pattern Analysis (MVPA) (see Section fNIRS analyses). Although cortical regions respond preferentially to particular stimuli categories, in infancy these responses are not as selective as in adults (Deen et al., 2017) and standard channel-by-channel analyses may not identify less stable or marked responses (Emberson et al., 2017). MVPA addresses this problem by asking whether it is possible to extract information about the experimental conditions from observed patterns of neural activations across multiple channels. This approach harnesses weakly discriminative information that is distributed over multiple channels and can therefore, in some cases, provide greater sensitivity than the univariate general linear model (Haynes & Rees, 2006). Recent work has successfully used MVPA to classify infant brain responses (Dopierała et al., 2023; Emberson et al., 2017; Mercure et al., 2019). In line with the pre-registered hypotheses, for these exploratory analyses we hypothesised that distributed patterns of cortical activity to visual speech and gurning would be classifiable at 10 months, but not at 5 months of age. Moreover, we hypothesised that as the left hemisphere becomes dominant for speech (Minagawa-Kawai et al., 2007), classifiable patterns would be observed over the left hemisphere only in the older age group.

2. Methods

2.1. Participants

An a-priori analysis (Dopierała et al., 2019), using G*Power software, assuming an alpha level of $\alpha = 0.05$, power of $1\text{-}\beta = 0.8$, and effect size of $\eta_p^2 = 0.24$ (e.g., Altvater-Mackensen & Grossmann, 2018; Lloyd-Fox et al., 2011), indicated a required sample size of 30 infants (15 per age group) for a repeated measures ANOVA, within-between interaction. The final sample consisted of 41 infants: 21 in the younger age group (5.0 – 6.6 months, M = 5.92, SD = 0.37) and 20 in the older age group (9.2 – 10.5 months, M = 9.89, SD = 0.43). All infants were born full-term (38–41 weeks), had normal birth weight (2790–4230 g, M = 3463, SD = 323.5), and Apgar scores above 5 (M = 9.6, SD = 1.2). Additional infants also participated but had to be excluded due to technical difficulties ($N_{older} = 2$), experimenter error ($N_{younger} = 1$, $N_{older} = 2$), improper headgear fitting or headgear moving during testing ($N_{younger} = 4$, $N_{older} = 11$), infant taking headgear off and/or pulling fibres out of the headgear during testing ($N_{older} = 2$), infant not sitting through the experiment (e.g., crying, moving a lot, $N_{younger} = 7$, $N_{older} = 19$), looking away from the screen ($N_{older} = 1$), exposure to a second language ($N_{younger} = 1$), data excluded during preprocessing ($N_{older} = 5$, see data exclusion criteria below), infant contributing less than 3 trials

per condition ($N_{younger}$ = 1, N_{older} = 6) or data from less than 31 channels (N_{older} = 2). Additionally, some infants were rescheduled and therefore had to be excluded for wrong age of testing ($N_{younger}$ = 7, N_{older} = 3). Although high, such attrition rate (64%) has been observed in infant fNIRS studies (Baek et al., 2021; Cristia et al., 2014). 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 session, all parents gave written informed consent. For their participation, the families received a diploma, a small gift (a baby book), and a video recording of their play in the laboratory.

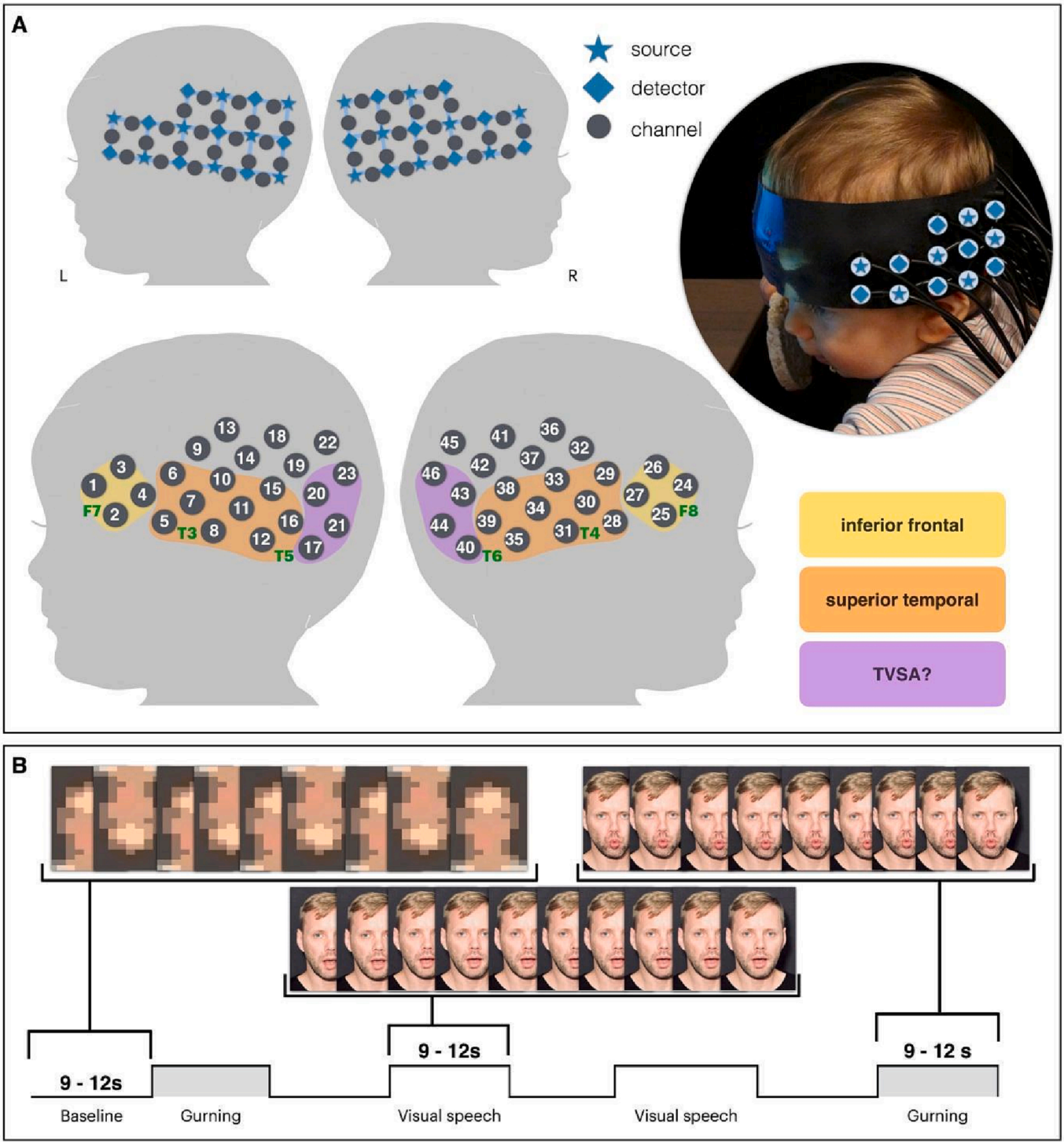

Fig. 1. FNIRS headgear and experimental paradigm. (A) Picture of an infant wearing the NTS fNIRS headgear used in the current study and illustrations of channel location in relation to infant's head: sources (stars) and detectors (diamonds), grey circles indicate measurement channels, the 10–20 coordinates superimposed on the diagram in green. Channels within yellow box are part of the inferior frontal region, orange - superior temporal region, purple - putative temporal visual speech area (TVSA). (B) Example stimuli sequence with still frames from experimental stimuli (for illustrative purposes, in the actual study stimuli included only female actresses). (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

2.2. Stimuli

Stimuli included video clips of two female native polish speakers either (1) articulating syllables (/ba/ or /ga/) or (2) gurning (pressing lips together, twirling closed mouth, moving mouth from side to side, puffing lips). The video clips showed a single speaker from the neck up on a dark grey background (see Fig. 1 B for illustrative purposes) looking directly at the camera. Both speech and gurning mouth movements lasted 1000 ms. Stimuli were edited into experimental trials: visual speech and gurning, which included 9 to 12 repetitions of the mouth movements (either syllable articulations or gurns). To increase infants' attention to the screen, trials included alternating stimuli of both speakers (random, 3 s to 4 s long intervals). Within a trial, the speakers were articulating the same syllable (either /ba/ or /ga/). The baseline was a visual stimulus containing non-social visual motion, created from static pixelated still frames from the experimental stimuli (edited with Movavi Video Editor software, version 15, Movavi, USA). For the baseline to elicit activation of visual motion processing regions, we edited the images to slowly zoom in, creating 3 s long videos. To make the baseline more engaging the baseline included alternate stimuli from a single speaker in a pseudorandom order (e.g., upright, inverted, upright). To control for low-level processing, we muted the auditory stream and set the average level of luminance the same for all stimuli. To avoid anticipatory brain activity and reduce the effect of physiological oscillations on optical signal (Lloyd-Fox et al., 2010; Peña et al., 2003), both experimental and baseline trials had jittered length: 9–12 s (e.g., Lloyd-Fox et al., 2009, 2011; Mercure et al., 2019). Trials were presented in a pseudo-random order, so that every two minutes 3 trials per experimental condition were presented (e.g., Lloyd-Fox et al., 2011).

2.3. Procedure

Infants sat on their parent's lap, approx. 60 cm from a screen, in a dimly lit room. We aligned the fNIRS headgear to the midline to the infant's nasion, and placing the sides so that the midpoint of the lower row of channels was above the pre-auricular points (Lloyd-Fox et al., 2009). We instructed parents to refrain from talking to or interacting with the baby throughout the procedure. To draw the infant's attention to the screen and away from the headgear being placed on their head, the experiment started with screen familiarisation (geometrical figures or a movie of an aquarium). Once the headgear was in place and the infant was looking at the screen, the experimental task started. On the screen, time-locked stimuli were presented using Psychtoolbox (Pelli, 1997) for MATLAB version 9.2 (R2017a, Mathworks Inc., Sherborn, MA, USA). The experimenter stood hidden from the infant, monitoring and recording their behaviour throughout the procedure, and manually triggering attention-grabbing sounds (occasional alerting sounds) when infants did not attend to the screen (random, once every few trials). Experiment ended when the infant became fussy or watched 18 experimental trials (9 per condition).

We recorded fNIRS data using an NTS optical topography system (Gowerlabs Ltd. L, UK) with two continuous wavelengths of source light: 780 and 850 nm. Infants wore a custom build CBCD fNIRS headgear (https://cbcd.bbk.ac.uk/node/165), consisting of two source-detector arrays (Fig. 1) with 46 channels (source-detector separations: 2 cm) stretching bilaterally from frontal to temporal lobes. Before the experiment, we measured infants' head circumference to determine headgear size. Previous infant MRI-NIRS co-registration study showed some of these channels to be sensitive to frontal, fronto-temporal, temporal, temporo-fronto-parietal, and temporoparietal regions (Lloyd-Fox et al., 2014). We extended posteriorly the array adding additional channels to cover approximately regions posterior and inferior to the pST, the likely region of the putative TVSA (Bernstein et al., 2011).

2.4. Analyses

Our fNIRS analysis plan was pre-registered (Dopierała et al., 2019). We pre-processed raw fNIRS data in HOMER2 (Huppert et al., 2009) following previously established pipelines and guidelines (Di Lorenzo et al., 2019). We excluded channels with raw intensities below 0.001 μM or above 10 μM or containing excessive motion artefacts (observed on over 3 trials throughout the testing session). We corrected motion artefacts with wavelet analyses (iqr = 0.8, Di Lorenzo et al., 2019) and spline correction (Molavi & Dumont, 2012; Scholkmann et al., 2010), which allows recovering most motion-affected trials compared to other motion correction methods (Brigadoi et al., 2014). We excluded trials which - post artefact correction - contained or were preceded (2 s) by significant motion artefacts and trials during which the infant looked away from the screen for over 60% of the time or parent interfered (e.g., talked to the baby). We excluded infants who contributed less than 3 trials per condition (considered enough to model the haemodynamic response in infants, Lloyd-Fox et al., 2010) or less than 31 channels (Mercure et al., 2019). We removed physiological noise using low-pass 0.50 Hz and high-pass 0.03 Hz filters (Di Lorenzo et al., 2019). Following this, we converted data to relative concentrations of HbO and HbR, assuming a differential pathway factor of 5.1 (e.g., Lloyd-Fox et al., 2010). Finally, we averaged data into 25 s blocks: 5 s pre-stimulus baseline and 20 s post-stimulus time period. We analysed the time period between 5 and 15 s post-stimulus onset, which was previously found to include the maximum range of observed HbO and HbR changes (e.g., Lloyd-Fox et al., 2011, 2015; Mercure et al., 2019). As the latency of peak HRF may vary in infants (e.g., Lloyd-Fox et al., 2010) and the analyses were pre-registered before exploring the data, in order to capture the peak response, we split the time period into two time windows: 5–10 s and 10–15 s post-stimulus (Lloyd-Fox et al., 2015, 2017). We found that the pre-registered time windows captured the peak change in both HbO and HbR, in both age groups (see Supplementary Material Figure S1). On average, the peak occurred between 6 and 13.7 s (M = 7.7 s) in the younger and 6–13.2 s (M = 9.9 s) in the older age group. We analysed the mean changes in the concentration of chromophores (e.g., Gervain et al., 2008). The raw data, the pre-processing script, and the pre-processed data with anonymised IDs are available at https://osf.io/sqjft/?view_only=41d20e906335497b9ad661d5f1fe2118.

To gain a fuller understanding of the development of specialisation of infant cortical representations and the processing network of visual speech, we used both pre-registered univariate analyses (Dopierała et al., 2019) and exploratory MVPAs. For univariate analyses, we used channel-by-channel ANOVAs (e.g., Peña et al., 2003). Firstly, as we expected an effect of age, we ran a three-way mixed ANOVA (age × condition × time) to compare responses to visual speech and gurning between the age groups. Specifically, we used the F-test with simple planned contrasts to investigate the effect of time on changes in mean concentration of HbO and HbR (μMol) over the three time windows: −5–0 s, 5–10 s, 10–15 s (e.g. (Lloyd-Fox et al., 2015). A significant increase in HbO or decrease in HbR indicated cortical activation (Meek, 2002). We also report channels showing inverted haemodynamic responses (HbO decrease and/or HbR increase) which may indicate processing difficulty, habituation, or developmentally transient response related to functional specialisation of the cortex (Issard & Gervain, 2018; Mercure et al., 2019). We do not report channels showing simultaneous changes of the same direction in both chromophores which likely reflect artefacts (Xu et al., 2017). We then followed this with a two-way mixed ANOVA (age × time) for visual speech and gurning separately to determine how age influenced responses to each condition. Secondly, we run a two-way RM ANOVA (condition × time window) to compare mean haemodynamic responses directly between experimental conditions (visual speech relative to gurning). Again, we used planned simple contrasts to identify selective channels, which show different responses (i.e., change in concentration of chromophores between baseline and either activation time window) depending on condition. Finally, we run

one-way RM ANOVA (time window) separately for each condition, expecting that different channels may be active during visual speech and gurning, however they may not show significantly different responses. We used planned simple contrasts to identify sensitive channels which show a significant difference in concentration of chromophores between the baseline and either activation time window. To account for the number of channels, we applied the FDR (Benjamini & Hochberg, 1995) correction for multiple comparisons separately for each channel-wise analysis. No results survived the correction (see Section 4) therefore we report uncorrected results. Additionally, in an exploratory analysis we correlated the response of channels identified as the likely region of the putative TVSA with age. We run separate correlations for responses to visual speech and gurning in each age group for both chromophores. All univariate analyses were conducted in SPSS version 27.0 (IBM Corp., NY, USA). The code to analyse the data in SPSS is available at https://osf.io/sqjft/?view_only=41d20e906335497b9ad661d5f1fe2118.

For the exploratory MVPAs, we followed the same method as Mercure et al. (2019). We used a Support Vector Machine (SVM) to learn a classification boundary that separates neural patterns associated with two labeled experimental conditions (Mercure et al., 2019). Once trained, the model can be tested by assessing its ability to successfully discriminate the conditions, for unseen data, in which the labels of the two conditions have been withheld. If the model is able to successfully predict the labels of the unseen data at a level greater than chance, we assume that the region contains sufficient information to discriminate between the experimental conditions (Haxby & Gobbini, 2012; Haxby et al., 2001).

A single neural pattern for each condition and participant was derived by averaging the neural response across all trials within each condition, for each activation time window (5–10 s, 10–15 s) and chromophore (HbO, HbR). The data were z-scored within each channel across all infants to ensure that the channels were on comparable scales for classification. We then used a leave-one-infant-out decoding approach (cf. Emberson et al., 2017), which is different to typical decoding approaches. To reduce the impact of noise in the data and "lost trials" - which are a feature of neural data acquired from infants - we generated a single pattern for each condition and participant by averaging data from all the trials for each condition and using this as input to the classification. Specifically, we generated a single *training* pattern for each condition by averaging the patterns across all participants, except the held-out participant, for each of the two experimental conditions. This trained model was then *tested* on its ability to predict the labels from the two patterns, one for each condition, from the held-out participant. This training and testing was done iteratively for all participants, leaving a different participant out each time until all participants had been selected as the test participant. An accuracy estimate for the group was then calculated by summing the number of correctly predicted patterns and dividing this by the total number of predictions to generate a proportion correct score. This provided a single accuracy score based on the correctly labeled data for the group of infants. Not all channels contained usable data for all participants. These channels were dropped out of both the training and test patterns when classifying participants that had missing channels. This meant that the exact channels used in the analysis differed slightly for each participant.

A classical parametric statistical approach was not appropriate for this kind of leave-one-infant-out cross validation approach, due to the violation of the i.i.d. assumption. Hence, we used a permutation based statistical approach. We compared the observed classification accuracy from the correctly labeled data to the classification accuracies arising from shuffling the condition labels and training and testing SVM models based on the shuffled data (Pereira et al., 2009). Note that this shuffling occurred at the level of the epoch averages – the averaged data submitted to classification – rather than by shuffling the labels of the individual trials and re-estimating the averages. The classifier was trained and tested 1000 times, generating 1000 classification accuracy scores in which for each of the 1000 permutations, the condition labels were shuffled for all participants. Note that both the training and test data were shuffled and the participant structure was maintained during the shuffling, such that the patterns representing each participant were randomly maintained or swapped with one another. This built a null-distribution that takes into account the dependencies and structure of the data that arise from the leave one infant out cross-validation approach using the correctly labeled data. Note that 1000 permutations were selected from a 2^n possible permutations (where n = number of participants). The probability value was established by ranking the observed accuracy for the group of infants using the correctly labeled data relative to the distribution of accuracies generated by training and testing on epoch averages in which the labels had been randomly shuffled. The observed accuracy score was included in both the numerator and denominator for calculating the p value, such that if the classification accuracy observed from the data was higher than all the observed permutation values, this would result in a value of $p = 1/1001$. This is a conservative approach, but was preferred as including the observed accuracy in both the numerator and denominator prevents an estimated p-value of zero in the instance that the observed value is the highest score attained (Ruxton & Neuhäuser, 2013).

For successful classifications, we report which channels contributed most to the classification (the classifier weight value for the channels), i.e., the informativeness of the channels, these are the weights/channels that have the greatest influence on the classification boundary. The weights for each channel were determined by re-training the classifier using the data averaged over all participants for each condition and extracting the weight vector of a model trained on these averaged patterns. To account for the fact that some channels were dropped out of the classification when calculating the classifier accuracy, due to missing channels, we trained this final model only using channels for which there was usable data in at least 80% of participants. The classification prediction (whether an unseen example is classified as belonging to one condition or the other), is achieved by summing the activation values at each channel multiplied by their associated weight value, and adding a bias term. Hence, rather than visualising the raw weight vector, we multiplied the weight vector by the average patterns to take into account the channel values, their associated weight and how combining these values influences the classification outcome. The most informative channels were defined as the channels contributing the 30% most extreme values. Note that due to the normalisation, the channels contributing most to classifying in favour of the positive class (e.g. one of the conditions) were the same as those contributing most to classifying the negative class (e.g. the other condition). As such, the weights reflected the channels that provided the most effective discrimination between conditions rather than necessarily characterising one condition or the other. As in univariate analyses, we conducted separate MVPAs on mean changes in HbO and HbR during each activation time window (5–10 and 10–15 s post-stimulus onset), and conducted separate analyses for each age group. To test for hemispheric contributions to classification, we conducted MVPAs separately on all, left, and right hemisphere channels. Multivariate analyses were conducted using a custom Matlab script (see https://github.com/speechAndBrains/fNIRS_tools).

3. Results

In both age group, infants contributed on average 4 visual speech and 4 gurning trials (younger age group: $M_{VS} = 4.62$, $SD_{VS} = 1.2$, $M_G = 4.48$, $SD_G = 0.9$, older age group: $M_{VS} = 4$, $SD_{VS} = 0.9$, $M_G = 4.05$, $SD_G = 1$). The difference in the number of trials contributed to each condition was not significant in either age group ($p > .5$).

3.1. Visual speech vs gurning

To assess which channels show differential response to visual speech versus gurning depending on age group (5-month-old vs 10-month-old), we ran 2x2x3 mixed ANOVA (age, condition, time [baseline vs T1,

baseline vs T2]). We used planned contrasts to identify channels showing a differential response (change between the baseline and either of the two activation time windows: 5–10 s, 10–15 s) depending on age and condition. Three superior temporal channels showed a significant interaction effect between the three variables: the left superior temporal channel (CH15, T1 $F(1,35) = 6.776$, $p = .014$) interaction reflected a higher HbO response to visual speech than gurning only in the younger but not the older age group, while the right superior temporal channels (CH34, T1 $F(1,30) = 5.615$, $p = .024$; CH39, T2 $F(1,27) = 5.911$, $p = .022$) interaction reflects a higher HbO and HbR response to gurning than visual speech only in the older age group. Two channels located superior to the left superior temporal region bilaterally (18, 36) showed significant interaction effects (CH18, T1 $F(1,25) = 4.305$, $p = .048$, T2 $F(1,25) = 4.374$, $p = .047$; CH36, T2 $F(1,25) = 5.659$, $p = .025$), reflecting a higher HbO response to visual speech than gurning only in the older age group. A right inferior frontal channel (CH27) showed significant interaction effects (HbO T1 $F(1,33) = 4.52$, $p = .039$, T2 $F(1,33) = 6.499$, $p = .016$; HbR T1 $F(1,33) = 4.294$, $p = .046$), reflecting a higher HbO response to gurning than visual speech only in the older age group, and a higher HbR response to visual speech than gurning only in the younger age group. A right superior temporal channel (CH34) showed significant interaction effect (T1 $F(1,30) = 5.615$, $p = .024$), reflecting a higher HbO response to gurning than visual speech only in the older age group. Finally, two channels – a left inferior frontal channel (CH4) and a channel located superior to the superior temporal region (CH41) – showed significant interaction effects (CH4, T1 $F(1,29) = 4.779$, $p = .037$, CH41, T1 $F(1,22) = 5.037$, $p = .035$). Results presented in Fig. 2 right panel; all statistics presented in inline Supplementary Table S4.

In the younger age group, the 2x3 RM ANOVA (condition × time [baseline vs T1, baseline vs T2]) with planned contrast revealed significant condition × time interactions suggesting that multiple channels showed significantly different responses depending on condition. A channel located superior to the left superior temporal region (CH19, T2 $p = .035$) showed significantly different HbR responses to visual speech than gurning. Other channels located over the right inferior frontal (CH27, T1 $p = .025$), bilateral superior temporal (CH11, CH28, CH38), and right TVSA homologue (CH43) regions showed different HbO (CH11, T1 $p = .044$, CH28, T2 $p = .027$) and HbR (CH27, T1 $p = .025$, CH38, T1 $p = .034$, CH43, T2 $p = .047$) responses depending on condition. Results presented in Fig. 2 left panel; all statistics presented in inline Supplementary Table S3.

In the older age group, the 2x3 RM ANOVA (condition × time [baseline vs T1, baseline vs T2]) with planned contrast revealed significant condition × time interactions also suggesting that multiple channels showed significantly different responses depending on condition. Channels located in the left putative TVSA (CH20, T1 $p = .045$), right inferior frontal (CH27, T2 $p = .032$) and superior temporal (CH34, T1 $p = .041$, T2 $p = .038$, CH39, T2 $p = .031$) regions showed differential responses. Additional channels located superior to the left (CH18, T1 $p = .043$, T2 $p = .02$) and right (CH32, T1 $p = .039$) superior temporal regions showed differential responses. Results presented in Fig. 2 middle panel; all statistics presented in inline Supplementary Table S3.

3.2. Visual speech vs non-social baseline

Between-group analyses – 2x3 (age × time [baseline vs T1, baseline vs T2]) mixed ANOVAs – revealed a significant age × time interaction effects suggesting that perception of visual speech elicited differential responses depending on age group. A right inferior frontal (CH24) channel showed different HbR responses in the older than younger age group (T2 $F(1,37) = 4.763$, $p = .036$). Post-hoc pairwise comparison showed that this difference reflected a greater HbR increase in the right inferior frontal region of older than younger infants ($p = .035$). Additionally, two channels (CH12, CH38) in the left and right superior temporal regions showed significant interaction effect of age and time (CH12, T2 $F(1,32) = 4.611$, $p = .049$; CH38, T1 $F(1,37) = 4.152$, $p = .039$), the left channel showed significantly greater HbO response in the younger than older age group ($p = .05$), while the right channel showed significantly greater HbR response in the older than younger age group ($p = .041$). However, post-hoc pairwise comparisons showed that none of the channels identified in the between-group analyses showed significant response (relative to baseline) to visual speech in either age group. Results presented in Fig. 3 top panel, all statistics presented in inline Supplementary Table S2.

In the younger age group, the planned contrast analyses using F-test (baseline vs T1, baseline vs T2, see Section 2.4) revealed significant effects. The perception of visual speech elicited significant increases in concentration of HbO over the left hemisphere: the superior temporal (CH10, T1 $p = .014$, T2 $p = .043$) and the putative TVSA (CH20, T2 $p = .004$; CH23, T1 $p = .02$). Additionally, a left frontal (CH3, T1 $p = .034$) channel showed an inverted response pattern: an increase in the concentration of HbR.

In the older age group, the planned contrast using F-test (baseline vs T1, baseline vs T2, see Section 2.4) revealed that visual speech did not elicit any significant increases in HbO concentration. Only one right superior temporal channel (CH30, T1 $p = .025$) showed activation, i.e., a decrease in HbR concentration. Other channels showed inverted responses: a right superior temporal (CH34, T1 $p = .009$) showed a decrease in concentration of HbO, right inferior frontal (CH24, T1 $p = .033$ and T2 $p = .041$, CH26, T1 $p = .026$) and superior temporal (CH35, T2 $p = .037$) showed an increase in concentration of HbR. Additional channels located superior to the superior temporal region showed

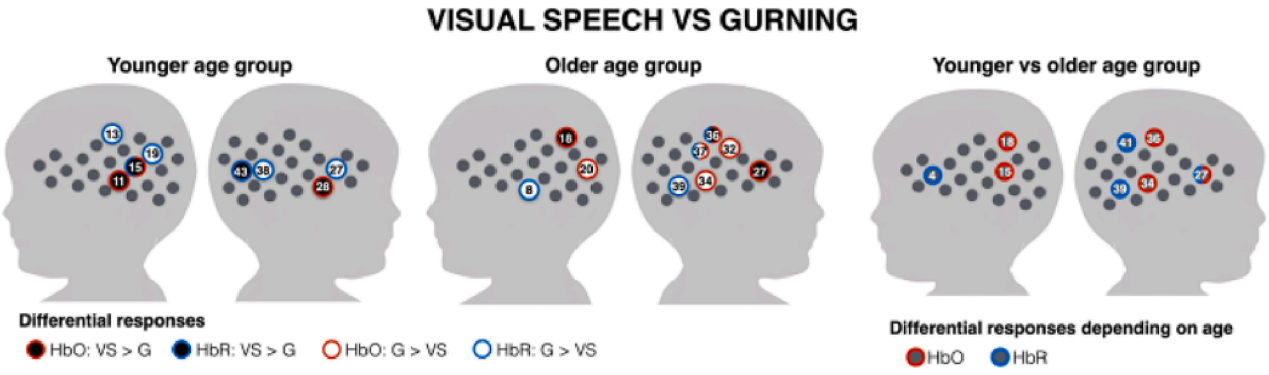

Fig. 2. Results of direct comparisons of visual speech versus gurning. Channels showing differential response to visual speech vs gurning in the younger (left column) and older (middle column) age groups, differential responses to visual speech vs gurning depending on age (right column) highlighted depending on the chromophore: red outline – HbO, blue outline – HbR, and direction of the effect: black – response to visual speech higher than gurning, white – response to gurning higher than visual speech. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

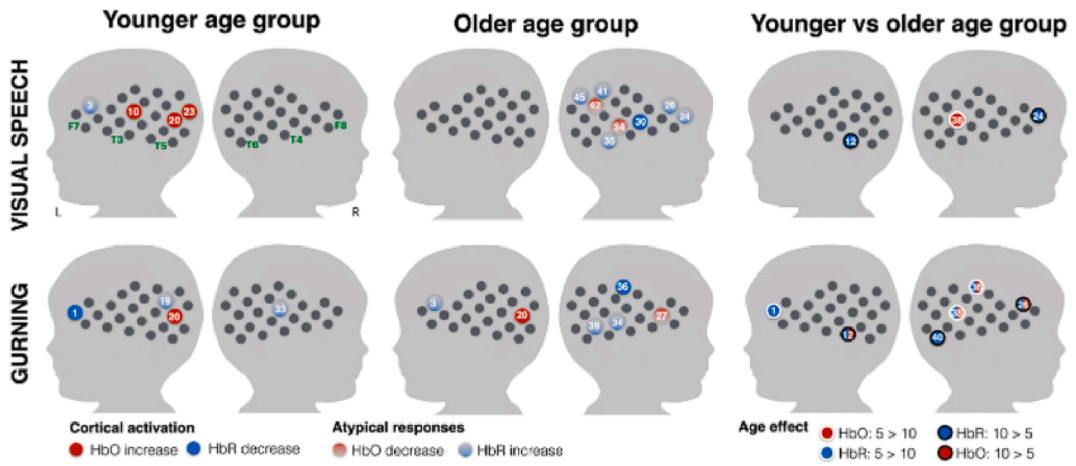

Fig. 3. Observed responses to visual speech (top row), and gurning (bottom row) in the younger (left column) and older (middle column) age groups, differential responses depending on age group (right column). Red: channels showing a significant increase in HbO in channel-by-channel analyses. Blue: channels showing a significant decrease in HbR. Light red: channels showing a significant decrease in HbO. Light blue: channels showing a significant increase in HbR. White outline: channels showing a greater response in the younger age group. Black outline: channels showing a greater response in the older age group. None of the results survived FDR correction for multiple comparisons. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

decreases in concentration of HbO (CH42, T1 p =.026) and increases in concentration of HbR (CH41, T1 p =.016, CH45, T2 p =.006). All statistics presented in inline Supplementary Table S1.

3.3. Gurning vs non-social baseline

Between-group analyses – 2x3 (age × time) mixed ANOVAs – revealed significant age × time interaction effects suggesting that perception of gurning elicited differential responses depending on age group. The left inferior frontal channel (CH1) showed a significantly different HbR response (T1 $F(1,38) = 6.88$, p =.012, T2 $F(1,28) = 4.365$, p =.043), reflecting a greater HbR decrease in the younger than older age group. A channel located superior to the superior temporal region (CH36) showed significantly different HbO and HbR responses (HbO T1 $F(1,26) = 7.97$, p =.009, HbR T1 $F(1,26) = 8.925$, p =.006), reflecting a greater HbR decrease in the older than younger age group. Additional four channels located in the left (CH12) and right (CH38) superior temporal regions, right inferior frontal (CH26), and right putative TVSA (CH40) regions showed significantly different HbO and HbR responses depending on age (CH12 HbO T1 $F(1,28) = 5.676$, p =.039, HbR T1 $F(1,28) = 2.422$, p =.009, CH 38 T2 HbO $F(1,36) = 8.498$, p =.006, HbR T2 $F(1,36) = 5.543$, p =.006, CH26 HbO T1 $F(1,37) = 10.536$, p =.002, HbR T1 $F(1,37) = 10.359$, p =.003, CH40 HbR T2 $F(1,31) = 8.93$, p =.005). For the right superior temporal channel (CH38) the effect reflected significantly greater responses in the younger than older age group (p =.008, p =.006). For the other three channels the effect reflected significantly greater responses in the older than younger age group (CH12 T1 HbO p =.038, HbR T1 p =.008, CH26 HbO T1 p =.002, HbR T1 p =.002, CH40 HbR T2 p =.005). However, planned contrasts (baseline vs T1 × age and baseline vs T2 × age) showed that none of the channels identified in between-group analyses showed significant response (relative to baseline) to gurning in either age group. Results presented in Fig. 3 bottom panel, all statistics presented in inline Supplementary Table S2.

In the younger age group, the planned contrast analyses using F-test (baseline vs T1, baseline vs T2, see Section 2.4) revealed that the perception of gurning elicited activation of two channels: The left putative TVSA (CH20, T2 p =.041) channel showed increases in HbO concentration, while the left inferior frontal (CH1, T1 p =.042) channel showed decreases in HbR concentration. Additionally, two channels showed inverted responses, i.e. increases in concentration of HbR: one channel superior to the left superior temporal region (CH19, T2 p =.005), and one right superior temporal channel (CH33, T1 p =.05).

In the older age group, the planned contrast analyses using F-test (baseline vs T1, baseline vs T2, see Section 2.4) revealed that the perception of gurning elicited bilateral activations: an increase in HbO concentration over the left putative TVSA channel (CH20, T1 p =.029) and a decrease in concentration of HbR in a channel superior to the right superior temporal region (CH36, T1 p =.047). It also led to inverted responses: a left inferior frontal (CH27, T2 p =.024) decrease in concentration of HbO, while left inferior frontal (CH3, T1 p =.003) and right superior temporal (CH34, T2 p =.049, CH39, T2p =.042) increases in HbR. All statistics presented in inline Supplementary Table S1.

3.4. Exploratory analyses

3.4.1. Correlation between TVSA response and age

To further explore the few significant activations (increase in HbO/decrease in HbR) to visual speech in the older age group, we correlated the mean HbO concentration during the activation time windows (5–10 s and 10–15 s) over the four TVSA channels (CH17, CH20, CH21, CH23) with age. We found a negative correlation within the first time window (5–10 s) over one TVSA channel on the left (CH20), $r(15) = -0.495$, p =.045 (not significant after FDR correction, see Fig. 4). Around 9 months of age visual speech elicited an HbO concentration increase, while around 10 months of age — a decrease. There was no correlation between the observed responses to gurning and age over the four TVSA channels (p >.05).

3.4.2. Multivariate pattern analyses

To assess differences in haemodynamic responses to visual speech and gurning on a network level, we conducted MVPAs for each age group, in each time window, for three sets of channels: all channels, left hemisphere channels, and right hemisphere channels, for both HbO and HbR chromophores (see Fig. 5). In the younger age group, MVPAs could not classify distributed patterns of HbO or HbR responses to visual speech and gurning at a level greater than chance in either time window, using any group of channels (all ps > 0.1; see Fig. 5 left column and Supplementary Table S5).

In the older age group, patterns of HbO responses to visual speech and gurning in the first time window (5–10 s) could be classified at a level greater than chance using all channels (proportion correct = 0.68, p =.02). The analysis revealed nine channels contributing most to the successful classification (see inline Supplementary Table S5 and Fig. 5, right column). These channels were located over the putative TVSA (CH20) and its right homologue region (CH40), left and right superior

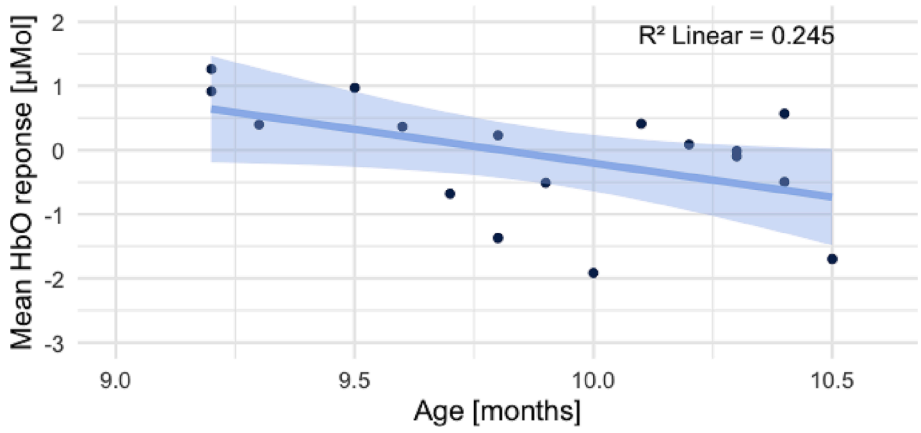

Fig. 4. Correlation of the mean HbO response (5–10 s) to visual speech and age observed in the putative TVSA (channel 20) in the older age group. Shaded area indicate 95% CI. The mean response became increasingly negative with age.

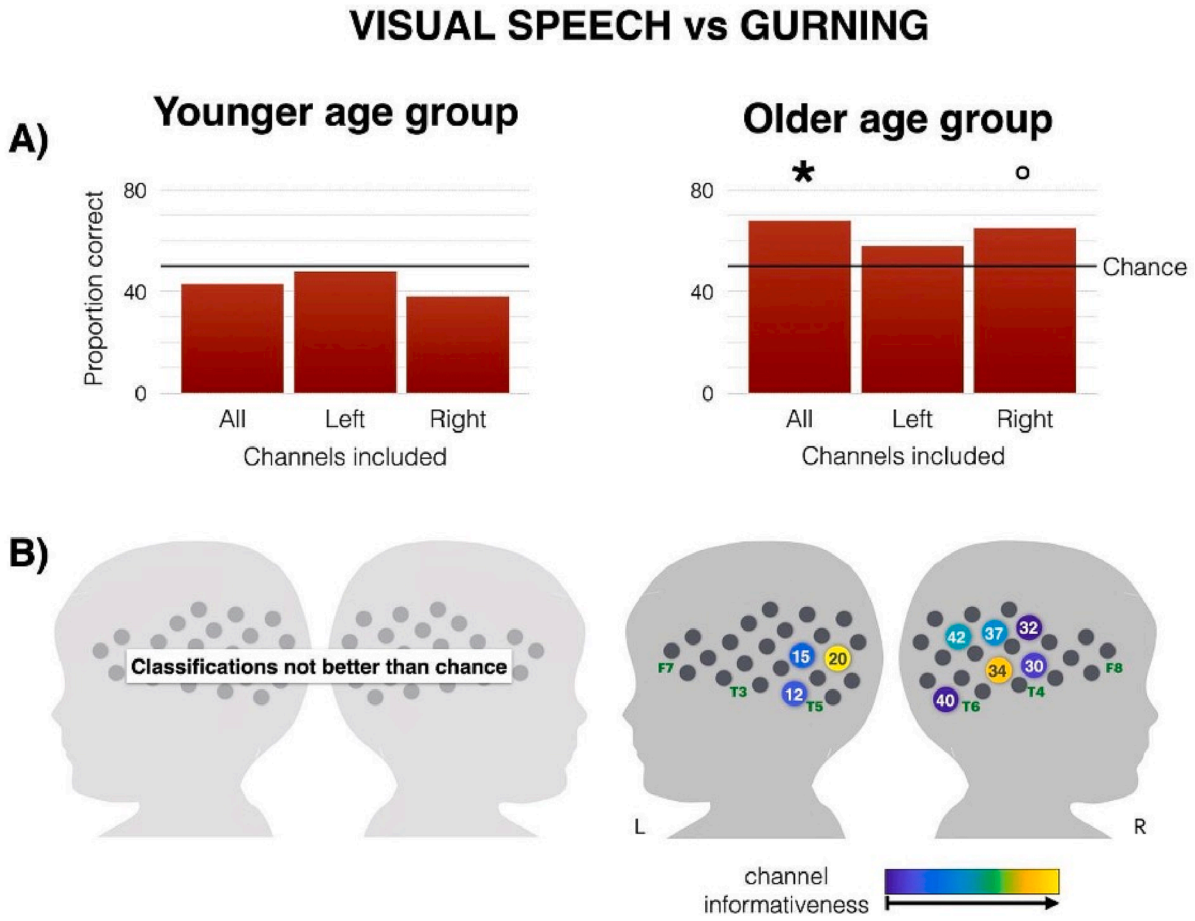

Fig. 5. Results of MVPA. Classification accuracy for visual speech versus gurning, presented for mean HbO concentration in the first time window (5–10 s). A) Observed proportion of correct classification for visual speech vs gurning depending on channel group: all – 46 channels, left – 23 channels located on the left hemisphere, right – 23 channels located on the right hemisphere. Asterix indicates significant result (p =.02), circle indicates trend result (p =.06). All channels contributed to successful classification. B) Graphical representation of the relative informativeness across channels from multivariate analysis (all channels were included, coloured channels that contributed the most). The 10–20 coordinates are superimposed on the diagram in green. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

temporal (CH12, CH15, CH30, CH34) regions, as well as right hemisphere channels located superior to the superior temporal region (32, 37, 42). Classifications of patterns of HbO responses using right hemisphere channels in the first time window (5–10 s) as well as all channels in the second time window (10–15 s) also reached high classification accuracies (65%, 65%) but only approached statistical significance (p =.055, p =.063). Specifically, we did not find any significant classifications using only left hemisphere channels. None of the other classifications of either HbO or HbR responses in either time window and channel group were significant (all ps > 0.09, see inline Supplementary Table S3).

4. Discussion

This cross-sectional study investigated the patterns of cortical responses to visual speech and gurning around 5 and 10 months of age. Using fNIRS, we measured cortical responses bilaterally over the inferior frontal and superior temporal regions, including the likely region of putative TVSA. By measuring such a wide area of the cortex across two age groups, we were able to show the development of regional sensitivity (relative to non-social motion) to visual speech and gurning in the time period when infants show increasing abilities to process faces and native language. In that, we extend findings from previous studies on speech processing which predominantly focused on measuring a single brain region (e.g., Fava et al., 2014) or age group (e.g., Altvater-Mackensen & Grossmann, 2018). By using a complex, dynamic baseline we were further able to dissociate activation to facial vs non-social motion, and directly compare the responses to different types of mouth movements. In that, we extend existing studies which predominantly used static (e.g., images of toys, Lloyd-Fox et al., 2009) or less specific (e.g., screensaver, Altvater-Mackensen & Grossmann, 2018) baseline stimuli. While preliminary, given that univariate channel-wise results did not survive the correction for multiple comparisons and some channels revealed inverted responses, our results suggest that visual speech elicits dissociable patterns of cortical activity (including the likely region of the putative TVSA) in 10-month-olds but not in 5-month-olds. These results potentially suggest that by 10 months of age infants are sensitive to the differences between visual speech and gurning. These findings imply that cortical representations of visual speech change between 5 and 10 months of age, but continue to develop beyond 10 months of age to become fully tuned to native language.

4.1. The development of cortical representations of visual speech

Taken together, our results suggest that cortical representations of visual speech become distinct from representations of other mouth movements only around 10 months of age. Around 5 months of age both visual speech and gurning elicited fronto-temporal activations (relative to the non-social dynamic baseline, uncorrected). No channels showed significantly different activation to the two types of mouth movements. Additionally, MVPA showed that on a network level, distributed patterns of cortical activity to visual speech and gurning were classified at chance level, which implies that distinct patterns of cortical responses to these mouth movements are not distinguishable at this age. We propose that around 5 months of age representations of visual speech are not yet specific to visible articulations of syllables, instead all types of mouth movements are represented similarly. Quinn et al., (2021) suggested that initially representations are categorised along perceptual differences. Our results imply that around 5 months of age cortical representations of visual speech are categorised by low-level visual features rather than higher-order visemic/phonological features.

Dissociable cortical representations of visual speech vs gurning emerged around 10 months of age. In particular, a single channel over the likely region of the putative TVSA (left hemisphere) was active to gurning relative to baseline and showed a significantly (uncorrected) higher response to gurning than visual speech. Additionally, MVPA correctly classified response patterns to visual speech and gurning, indicating that visual speech elicits differential distributed patterns of cortical responses compared to gurning. The classification accuracy of the MVPA reached 68%, which is reasonably high and comparable to previous infant fNIRS (72% Emberson et al., 2017; 68% Mercure et al., 2019) and adult fMRI studies (e.g., Evans et al., 2014; McGettigan et al., 2012; Misaki, et al., 2010). Even though we used infant-level decoding (average pattern across all trials), the accuracy was similar as for MVPAs which used trial-level decoding (average pattern for a single trial). This result is all the more remarkable given that infants contribute data of lower quality (more motion artefacts) and quantity (fewer channels, fewer trials per condition, greater inter-subject variability) than adults (Emberson et al., 2017). The finding that around 10 months of age infants have different cortical responses to different mouth movements (visual speech vs gurning) is consistent with a previous study reporting differential temporal responses to familiar and unknown non-speech mouth movements in 5- and 8-month-olds (Tsurumi et al., 2019). We extend this finding by showing that cortical representations of visual speech do not become specific to visible articulations of syllables until after 9–10 months of age. These results imply that cortical representations of visual speech begin as general mouth movements detectors, that only later become specific to visible articulatory mouth movements.

Interestingly, while within-age-group analyses revealed different patterns of responses in younger and older infants to both visual speech and gurning, the between-age-group analyses identified few channels that show significantly different responses to the two conditions. Although visual speech elicited significant cortical activations only in the younger age group but not in the older, we did not find a significant age effect on activations in univariate analyses. The three channels that showed an effect of age showed either an atypical response (i.e., HbR increase) or no significant change relative to baseline. We speculate that the lack of significant age effect is related to the ongoing process of specialisation of the putative TVSA around 10 months of age. The observed correlation between the likely region of the putative TVSA's responses to visual speech and age in the older age group suggests that cortical representations of visual speech become re-organised between 9 and 10 months of age. Possibly 9-month-olds still represented visual speech as any mouth movement, showing an increased response in the likely region of the putative TVSA to visual speech (as in the younger age group). On the other hand, 10-month-olds likely started to represent visual speech as speech, showing a negative TVSA response (due to processing difficulty, Issard & Gervain, 2018). Interestingly, the correlation between age and TVSA activation was only significant for responses to visual speech, not for gurning, indicating that during this short period of time cortical specialisation for speech begins to emerge. Moreover, we expected the putative TVSA to show increasingly different responses to visual speech versus gurning with age. While the activation in the likely region of the putative TVSA was significantly different to gurning than to visual speech only in the older age group, between-age group analyses showed that this effect was not significantly different depending on the age group. Together, these results imply that infants' visual speech processing network is still developing at ten months of age, and 9 to 10 months of age is a transition period during which the network becomes re-organised.

The observed changes in the specialisation of the likely region of the putative TVSA are likely related to changes in infants' attention to audiovisual speech. Infants' looking patterns to audiovisual speech change within the first year of life: with age, infants look increasingly towards the mouth of the speaker (Lewkowicz & Hansen-Tift, 2012; Lozano et al., 2022; Mercure et al., 2019; Tomalski et al., 2013). This increased looking time to the mouth likely reflects increased attention to multisensory cues and/or using visual speech to aid auditory speech processing. Based on the combined results of previous eye-tracking studies and our fNIRS study, we can expect that the specialisation of TVSA continues to develop further as infants grow and gain more experience with language. Infants may become more sensitive to the subtle differences between different speech sounds, and may develop more refined representations of these sounds in the brain. Recent studies on audiovisual speech processing show that the trajectory of development of attention to mouth of talking faces differs depending on language and/or type of language (spoken vs sign) (Mercure et al, 2019; Lozano et al., 2022). As infants gain more exposure to different languages and dialects, their TVSA may become more specialized for the specific characteristics of those languages. It is also possible that the specialization of TVSA may continue to change throughout development, as children learn to use visual speech information in more sophisticated ways. For example, children may learn to integrate visual and auditory speech information more effectively, or may learn to selectively attend to visual speech

information in noisy or challenging listening environments. Phonological development continues well into the second year of life and beyond, so we would expect that the TVSA specialises further as phoneme/viseme categories improve. Overall, it is clear that the development of TVSA is a complex process that involves a variety of factors, including experience with language, attention, and perceptual processing. Further research is needed to fully understand how this process unfolds over time, and how it contributes to the development of language and communication skills.

Given that the observed cortical responses were initially selective to non-specific facial movements (gurning) rather than visual speech, the current study provides preliminary evidence that the putative TVSA initially develops as part of the dynamic face processing network rather than the speech network. Further analysis of the MVPA channel weights revealed that the channel in the left likely region of the putative TVSA contributed most to the successful classification in the older age group. The channel in the right likely region of the putative TVSA was also among the most informative channels. In adults, the putative TVSA represents visual syllables (Bernstein & Liebenthal, 2014): it shows greater responses to visual speech than gurning. By contrast, bilateral posterior superior temporal regions (the putative TVSA and right TVSA homologue) represent any type of mouth movements (Files et al., 2013). In our study, the observed pattern of responses of the left likely region of the putative TVSA was consistent with responses of the right TVSA homologue in adults (Files et al., 2013). Increased neural specialisation leads to emergence of response-selective tissue with age (Johnson, 2001; 2011). Thus, the selective responses to gurning observed in the older but not younger infants could potentially reflect increasing specialisation of the cortical network supporting the processing of dynamic mouth movements (Johnson, 2001; 2011).

4.2. Limitations and future directions

Presented results should, however, be interpreted with some caution, as we note some limitations. First, none of the ANOVAs results of individual channel activations were significant after the FDR correction for multiple comparisons (Benjamini & Hochberg, 1995). With every additional channel included in the analyses, the correction requires a smaller p-value for results to remain significant. The present study used a large headgear, covering not only bilateral fronto-temporal cortices but also regions located superior and posterior to the superior temporal lobe. In that we were able to show that predominantly the fronto-temporal cortices and the likely region of the putative TVSA were selectively responsive to visual speech and gurning. We argue that the univariate approach - which requires correction for multiple comparisons - is a suboptimal method for analysing multi-channel, developmental fNIRS data. A multivariate approach, such as the MVPA, is increasingly being used to analyse developmental fNIRS data (e.g., Emberson et al., 2017; Mercure et al., 2019).

Moreover, very few channels showed uncorrected significant activation to presented stimuli, and some channels showed inverted responses (decrease in HbO or increase in HbR). Few observed responses could be related to the employed baseline. A dynamic baseline - although rarely used in previous infant studies - allowed us to control for low-level visual and motion processing and to draw stronger conclusions regarding the sensitivity of cortical regions. Initially, the infant cortex is not as selective as in adults. Regions sensitive to dynamic social stimuli are also likely active to dynamic non-social stimuli. Relative to a mechanical motion condition, Lloyd Fox et al., (2011) observed few channels where HbO increased to mouth movements. Likewise, in our study we also observed few activations relative to the presented dynamic baseline.

The inverted responses are not uncommon in infants (e.g., Issard & Gervain, 2018; Kobayashi et al., 2011) and might be observed for a number of reasons. Repetition suppression – a phenomenon where repetition of a stimulus results in a decreased cortical response (Grill-Spector et al., 1999) – may explain the inverted responses. While repetition of stimuli is commonly employed and would therefore contribute to observed patterns of activation in most infant fNIRS studies, our task design may be particularly prone to its effects as the repetition differed between the two conditions. While in the visual speech condition infants observed two types of mouth movements (syllables /ba/ and /ga/, 2 from each speaker), in the gurning condition they observed a total of 4 different types of mouth movement (2 from each speaker). However, given that we observed inverted responses to both visual speech and gurning, we find this interpretation unlikely. On the other hand, inverted responses may result from immature vascular coupling (Lloyd-Fox et al., 2010), greater metabolic demands (Meek, 2002), or reflect transient developmental responses (Mercure et al., 2019). Moreover, too complex or demanding (or too simple) stimuli often elicit inverted responses (Issard & Gervain, 2018). Finally, the dynamic baseline may have contributed to the observed inverted responses, but more research is needed to understand the effects of different types of baselines on cortical responses in infants.

Secondly, the relatively low sample size suggests a need for replication. Such sample size (approx. 20 infants per group) is typical of infant fNIRS studies, as the population is difficult to test and attrition rates are high (Baek et al., 2021). The fact that the task was silent (and thus less engaging for infants) and the large headgear used (46 channels), likely further contributed to high attrition (65%). While driving attrition, a silent task was optimal to describe the development of the visual speech network: Conclusions from a non-silent task would be limited, as background noise impacts perception of visual speech (Nath & Beauchamp, 2011) and auditory processing elicits partially overlapping patterns of cortical activation with visual speech (e.g., Pekkola et al., 2005). A lower number of channels would likely limit attrition (Baek et al., 2021) but would not allow us to make inferences regarding the development of lateralisation (for unilateral arrays, e.g., Lloyd-Fox et al., 2017) and/or development of sensitivity/selective of wide areas of the cortex (for arrays only covering a single lobe, e.g., Tsurumi et al., 2020). Additionally, future studies may employ a longitudinal –rather than cross-sectional – design to examine individual trajectories of the development of brain specialisation for visual speech.

Thirdly, the location of observed activations has to be considered with some caution. According to previous co-registrations, our headgear covered inferior frontal, superior temporal (Lloyd-Fox et al., 2014), and the middle temporal (Perdue et al., 2019) regions bilaterally. However, given that fNIRS has lower spatial resolution than fMRI, our results cannot differentiate between the anterior and posterior parts of the pSTS, which in adults shows functionally different responses to during vs visual speech (Venezia et al., 2017). Similarly, the spatial resolution of fNIRS is not high enough to distinguish between the putative TVSA and the V5/MT. According to the Bernstein and Liebenthal model (2014), the TVSA is located between the pSTS and the motion selective V5/MT area. The V5/MT would likely respond to any motion, so lack of a response to visual speech leads us to believe that we did measure the likely region of the putative TVSA. Future studies could employ additional methods to visualise the location of the headgear (e.g., Jaffe-Dax et al., 2020), reconstruct the image (e.g., Zhao et al., 2021), or use fMRI with infants (e.g., Dehaene-Lambertz, et al. 2002) to increase our understanding of spatial specificity of the observed results.

Finally, the observed predominance of higher responses to gurning than visual speech might reflect task design rather than developmental effects. Given the higher rate of stimuli repetition, it is possible that the visual speech condition elicited more repetition suppression effects than the gurning condition. Further studies could clarify this by presenting more varied visual movements in each condition or limiting the type of gurning mouth movements to match the visual speech movements.

5. Conclusions

This is the first study to cross-sectionally assess age differences in how infants process silent visual syllables at a cortical level. By comparing cortical responses to visual speech and gurning with responses to non-social dynamic stimuli (baseline), we identified cortical regions involved in processing dynamic faces. Additionally, we identified neural responses that were specific to visual speech by comparing responses to visual speech with non-communicative mouth movements. Although preliminary, given their uncorrected and partially non-standard nature (i.e., observed both standard and inverted hemodynamic responses), our results offer insight into the development of functional cortical specialisation for visual speech in infancy. Particularly, around five months of age, the putative Temporal Visual Speech Area (TVSA) seems already sensitive to both visual speech and gurning, but it does not show distinct responses to the two conditions until 10 months of age. These findings suggest that initially, cortical representations of visual speech are non-differentiable from representations of other mouth movements. However, these representations become specific to discriminable mouth movements rather than visual speech towards the end of the first year of life. Altogether, these results provide further evidence of the existence of the putative TVSA as the site of cortical representations of visemes and imply that the TVSA starts to specialise in late infancy.

Data and code availability statement: The paper uses data collected at the University of Warsaw, Poland. The anonymised raw and pre-processed data, the script to pre-process the data in HoMer, and the code to analyse the data in SPSS are all available at https://osf.io/sqjft/?view_only=41d20e906335497b9ad661d5f1fe2118. The custom Matlab code used to run MVPA analyses is available at https://github.com/speechAndBrains/fNIRS_tools.

Funding sources: 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 analyses was provided by the Institute of Psychology, PAS.

Ethics approval statement: 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 session, all parents gave written informed consent.

CRediT authorship contribution statement

Aleksandra A.W. Dopierała: Conceptualization, Methodology, Formal analysis, Investigation, Writing – original draft, Writing – review & editing, Visualization. **David López Pérez:** Software, Methodology, Writing – review & editing. **Evelyne Mercure:** Conceptualization, Writing – review & editing, Supervision. **Agnieszka Pluta:** Conceptualization, Writing – review & editing, Supervision. **Anna Malinowska-Korczak:** Conceptualization, Investigation. **Samuel Evans:** Software, Writing – review & editing. **Tomasz Wolak:** Conceptualization. **Przemysław Tomalski:** Conceptualization, Methodology, Resources, Supervision, Project administration, Funding acquisition, Writing – review & editing.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

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. We also thank Dr. Danijel Korzinek and the late Prof. Krzysztof Marasek from the Polish-Japanese Academy of Information Technology for providing facilities to record and prepare experimental stimuli. Finally, we thank Zuzanna Laudańska, Magdalena Szmytke, Dianna Ilyka and the whole Babylab team for their help with data collection. 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 analyses was provided by the Institute of Psychology, PAS.

Appendix A. Supplementary material

Supplementary data to this article can be found online at https://doi.org/10.1016/j.bandl.2023.105304.

References

Altvater-Mackensen, N., & Grossmann, T. (2018). Modality-independent recruitment of inferior frontal cortex during speech processing in human infants. *Developmental Cognitive Neuroscience, 34*(August), 130–138. https://doi.org/10.1016/j.dcn.2018.10.002

Baek, S., Marques, S., Casey, K., Testerman, M., McGill, F., & Emberson, L. (2021). Attrition Rate in Infant fNIRS Research: A Meta-Analysis. *bioRxiv*.

Benjamini, Y., & Hochberg, Y. (1995). Controlling the false discovery rate: A practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society: Series B (Methodological), 57*(1). https://doi.org/10.1111/j.2517-6161.1995.tb02031.x

Bernstein, L. E., Jiang, J., Pantazis, D., Lu, Z. L., & Joshi, A. (2011). Visual phonetic processing localized using speech and nonspeech face gestures in video and point-light displays. *Human Brain Mapping, 32*(10), 1660–1676. https://doi.org/10.1002/hbm.21139

Bernstein, L. E., & Liebenthal, E. (2014). Neural pathways for visual speech perception. *Frontiers in Neuroscience*, *8*(DEC), 1-18. 25520611. doi: 10.3389/fnins.2014.00386 Brainard, D. H. (1997). The Psychophysics Toolbox. *Spatial Vision*, *10*(4). doi: 10.1163/156856897X00357.

Brigadoi, S., Ceccherini, L., Cutini, S., Scarpa, F., Scatturin, P., Selb, J., … Cooper, R. J. (2014). Motion artifacts in functional near-infrared spectroscopy: A comparison of motion correction techniques applied to real cognitive data. *NeuroImage, 85*(181–191), 23639260. https://doi.org/10.1016/j.neuroimage.2013.04.082

Campbell, R., MacSweeney, M., Surguladze, S., Calvert, G. A., McGuire, P., Suckling, J., … David, A. S. (2001). Cortical substrates for the perception of face actions: An fMRI study of the specificity of activation for seen speech and for meaningless lower-face acts (gurning). *Brain research. Cognitive brain research, 12*(2), 233–243.

Cristia, A., Minagawa-Kawai, Y., Egorova, N., Gervain, J., Filippin, L., Cabrol, D., & Dupoux, E. (2014). Neural correlates of infant accent discrimination: An fNIRS study. *Developmental Science, 17*(4), 628–635. https://doi.org/10.1111/desc.12160

Dehaene-Lambertz, G., Dehaene, S., & Hertz-Pannier, L. (2002). Functional neuroimaging of speech perception in infants. *Science, 298*(5600), 2013–2015.

Deen, B., Richardson, H., Dilks, D. D., Takahashi, A., Keil, B., Wald, L. L., … Saxe, R. (2017). Organization of high-level visual cortex in human infants. *Nature Communications, 8*. https://doi.org/10.1038/ncomms13995

Dick, A. S., Solodkin, A., & Small, S. L. (2010). Neural development of networks for audiovisual speech comprehension. *Brain and language, 114*(2), 101–114. https://doi.org/10.1016/j.bandl.2009.08.005.Neural

Dopierała, A. W., López Pérez, D., Mercure, E., Pluta, A., Wolak, T., & Tomalski, P. (2019, February 18). Development of Cortical Responses to Visual speech in Infancy. Retrieved from https://osf.io/sqjft/?view_only=41d20e906335497b9ad661d5f1fe2118.

Dopierała, A. A., Pérez, D. L., Mercure, E., Pluta, A., Malinowska-Korczak, A., Evans, S., … Tomalski, P. (2023). The development of cortical responses to the integration of audiovisual speech in infancy. *Brain Topography*. https://doi.org/10.1007/s10548-023-00959-8

Emberson, L. L., Zinszer, B. D., Raizada, R. D. S., & Aslin, R. N. (2017). Decoding the infant mind: Multivariate pattern analysis (MVPA) using fNIRS. *PLoS ONE, 12*(4), 1–23. https://doi.org/10.1371/journal.pone.0172500

Evans, S., Kyong, J. S., Rosen, S., Golestani, N., Warren, J. E., McGettigan, C., … Scott, S. K. (2014). The pathways for intelligible speech: Multivariate and univariate perspectives. *Cerebral Cortex, 24*(9), 2350–2361. https://doi.org/10.1093/cercor/bht083

Fava, E., Hull, R., & Bortfeld, H. (2014). Dissociating cortical activity during processing of native and non-native audiovisual speech from early to late infancy. *Brain Sciences, 4*(3), 471–487. https://doi.org/10.3390/brainsci4030471

Files, B. T., Auer, E. T., & Bernstein, L. E. (2013). The visual mismatch negativity elicited with visual speech stimuli. *Frontiers in Human Neuroscience, 7*(JUN), 371. https://doi.org/10.3389/fnhum.2013.00371

Gervain, J., Macagno, F., Cogoi, S., Pena, M., & Mehler, J. (2008). The neonate brain detects speech structure. *Proceedings of the National Academy of Sciences, 105*(37), 14222–14227. https://doi.org/10.1073/pnas.0806530105

Grossmann, T., Johnson, M. H., Lloyd-Fox, S., Blasi, A., Deligianni, F., Elwell, C., & Csibra, G. (2008). Early cortical specialization for face-to-face communication in human infants. *Proceedings of the Royal Society B: Biological Sciences*, *275*(1653), 2803-2811. 18755668. doi: 10.1098/rspb.2008.0986.

Hall, D. A., Fussell, C., & Summerfield, A. Q. (2005). Reading fluent speech from talking faces: Typical brain networks and individual differences. *Journal of Cognitive Neuroscience, 17*(6), 939–953. https://doi.org/10.1162/0898929054021175

Haxby, J. V., & Gobbini, M. I. (2012). Distributed neural systems for face perception. *Oxford Handbook of Face Perception, 93–110*. https://doi.org/10.1093/oxfordhb/9780199559053.013.0006

Haxby, J. V., Gobbini, M. I., Furey, M. L., Ishai, A., Schouten, J. L., & Pietrini, P. (2001). Distributed and overlapping representations of faces and objects in ventral temporal cortex. *Science, 293*(5539). https://doi.org/10.1126/science.1063736

Haynes, J. D., & Rees, G. (2006). *Decoding mental states from brain activity in humans* (Vol. 7). https://doi.org/10.1038/nrn1931

Huppert, T. J., Diamond, S. G., Franceschini, M. A., & Boas, D. A. (2009). HomER: A review of time-series analysis methods for near-infrared spectroscopy of the brain. *Applied Optics, 48*, 19340120. https://doi.org/10.1364/AO.48.00D280

Issard, C., & Gervain, J. (2018). Variability of the hemodynamic response in infants: Influence of experimental design and stimulus complexity. *Developmental Cognitive Neuroscience*, *33*(February 2017), 182-193. doi: 10.1016/j.dcn.2018.01.009.

Jaffe-Dax, S., Bermano, A. H., Erel, Y., & Emberson, L. L. (2020). Video-based motion-resilient reconstruction of three-dimensional position for functional near-infrared spectroscopy and electroencephalography head mounted probes. *Neurophotonics, 7*(03), 1. https://doi.org/10.1117/1.nph.7.3.035001

Johnson, M. H. (2001). Functional brain development in humans. *Nature Reviews Neuroscience, 2*, 11433372. https://doi.org/10.1038/35081509

Johnson, M. H. (2011). Interactive Specialization: A domain-general framework for human functional brain development? *Developmental Cognitive Neuroscience, 1(1), 7–21*, 22436416. https://doi.org/10.1016/j.dcn.2010.07.003

Kobayashi, M., Otsuka, Y., Nakato, E., Kanazawa, S., Yamaguchi, M. K., & Kakigi, R. (2011). Do infants represent the face in a viewpoint-invariant manner? Neural adaptation study as measured by near-infrared spectroscopy. *Frontiers in Human Neuroscience, 5*, 153.

Kubicek, C., de Boisferon, A. H., Dupierrix, E., Pascalis, O., Lœvenbruck, H., Gervain, J., & Schwarzer, G. (2014). Cross-modal matching of audio-visual German and French fluent speech in infancy. *PLoS ONE, 9*(2), e89275.

Kuhl, P. K., & Meltzoff, A. N. (1982). The bimodal perception of speech in infancy. *Science, 218*(4577), 1138–1141. https://doi.org/10.1126/science.7146899

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.

Lloyd-Fox, S., Begus, K., Halliday, D., Pirazzoli, L., Blasi, A., Papademetriou, M., … Elwell, C. E. (2017). Cortical specialisation to social stimuli from the first days to the second year of life: A rural Gambian cohort. *Developmental Cognitive Neuroscience, 25*, 92–104. https://doi.org/10.1016/j.dcn.2016.11.005

Lloyd-Fox, S., Blasi, A., & Elwell, C. E. (2010). Illuminating the developing brain: The past, present and future of functional near infrared spectroscopy. *Neuroscience and Biobehavioral Reviews*, *34*(3), 269-284. 19632270. doi: 10.1016/j.neubiorev.2009.07.008.

Lloyd-Fox, S., Blasi, A., Everdell, N., Elwell, C. E., & Johnson, M. H. (2011). Selective cortical mapping of biological motion processing in young infants. *Journal of Cognitive Neuroscience, 23*(9), 2521–2532. https://doi.org/10.1162/jocn.2010.21598

Lloyd-Fox, S., Blasi, A., Volein, A., Everdell, N., Elwell, C. E., & Johnson, M. H. (2009). Social perception in infancy: A near infrared spectroscopy study. *Child Development, 80*(4), 986–999. https://doi.org/10.1111/j.1467-8624.2009.01312.x

Lloyd-Fox, S., Richards, J. E., Blasi, A., Murphy, D. G. M., Elwell, C. E., & Johnson, M. H. (2014). Coregistering functional near-infrared spectroscopy with underlying cortical areas in infants. *Neurophotonics, 1*(2), Article 025006. https://doi.org/10.1117/1.nph.1.2.025006

Lloyd-Fox, S., Széplaki-Köllod, B., Yin, J., & Csibra, G. (2015). Are you talking to me? Neural activations in 6-month-old infants in response to being addressed during natural interactions. *Cortex, 70*, 35–48. https://doi.org/10.1016/j.cortex.2015.02.005

Lorenzo, R. D., Pirazzoli, L., Blasi, A., Bulgarelli, C., Hakuno, Y., Minagawa, Y., & Brigadoi, S. (2019). Recommendations for motion correction of infant fNIRS data applicable to data sets acquired with a variety of experimental designs and acquisition systems. *NeuroImage, 200*(June), 511–527. https://doi.org/10.1016/j.neuroimage.2019.06.056

MacKain, K., Studdert-Kennedy, M., Spieker, S., & Stern, D. (1983). Infant Intermodal Speech Perception Is a Left-Hemisphere Function. *Science, 219*(4590), 1347–1349.

Matchin, W., Groulx, K., & Hickok, G. (2014). Audiovisual speech integration does not rely on the motor system: Evidence from articulatory suppression, the McGurk Effect, and fMRI. *Journal of Cognitive Neuroscience, 26*(3), 606–620. https://doi.org/10.1162/jocn_a_00515

McGettigan, C., Evans, S., Rosen, S., Agnew, Z. K., Shah, P., & Scott, S. K. (2012). An Application of univariate and multivariate approaches in fMRI to quantifying the hemispheric lateralization of acoustic and linguistic processes. *Journal of Cognitive Neuroscience, 24*(3), 636–652. https://doi.org/10.1162/jocn_a_00161

Meek, J. (2002). Basic principles of optical imaging and application to the study of infant development. *Developmental Science, 5*(3), 371–380. https://doi.org/10.1111/1467-7687.00376

Mercure, E., Evans, S., Pirazzoli, L., Goldberg, L., Bowden-Howl, H., Coulson-Thaker, K., … MacSweeney, M. (2019). Language experience impacts brain activation for spoken and signed language in infancy: Insights from unimodal and bimodal bilinguals. *Neurobiology of Language, 1*(1), 1–24. https://doi.org/10.1162/nol_a_00001

Minagawa-Kawai, Y., Mori, K., Naoi, N., & Kojima, S. (2007). Neural attunement processes in infants during the acquisition of a language-specific phonemic contrast. *Journal of Neuroscience, 27*(2), 315–321.

Misaki, M., Kim, Y., Bandettini, P. A., & Kriegeskorte, N. (2010). Comparison of multivariate classifiers and response normalizations for pattern-information fMRI. *NeuroImage, 53*(1), 103–118. https://doi.org/10.1016/j.neuroimage.2010.05.051

Molavi, B., & Dumont, G. A. (2012). Wavelet-based motion artifact removal for functional near-infrared spectroscopy. *Physiological Measurement, 33*(2), 259–270. https://doi.org/10.1088/0967-3334/33/2/259

Ojanen, V. (2005). *Neurocognitive mechanisms of audiovisual speech perception.* doi: 10.3389/conf.neuro.01.2009.04.090.

Okada, K., & Hickok, G. (2009). Two cortical mechanisms support the integration of visual and auditory speech: A hypothesis and preliminary data. *Neuroscience Letters, 452*(3), 219–223. https://doi.org/10.1016/j.neulet.2009.01.060

Olson, I. R., Gatenby, J. C., & Gore, J. C. (2002). A comparison of bound and unbound audio-visual information processing in the human cerebral cortex. *Cognitive Brain Research, 14*(1), 129–138. https://doi.org/10.1016/S0926-6410(02)00067-8

Patterson, M. L., & Werker, J. F. (1999). Matching phonetic information in lips and voice is robust in 4.5-month-old infants. *Infant Behavior and Development, 22*(2), 237–247. https://doi.org/10.1016/S0163-6383(99)00003-X

Patterson, M. L., & Werker, J. F. (2003). Two-month-old infants match phonetic information in lips and voice. *Developmental Science, 6*(2), 191–196. https://doi.org/10.1111/1467-7687.00271

Pekkola, J., Ojanen, V., Autti, T., Jääskeläinen, I., Möttönen, R., Tarkiainen, R., & Sams, M. (2005). Primary auditory cortex activation by visual speech: An fMRI study at 3 T. *Neuroreport, 16*(2), 125–128.

Pelli, D. G. (1997). The VideoToolbox software for visual psychophysics: Transforming numbers into movies. *Spatial Vision, 10*(4). https://doi.org/10.1163/156856897X00366

Perdue, K. L., Jensen, S. K., Kumar, S., Richards, J. E., Kakon, S. H., Haque, R., … Nelson, C. A. (2019). Using functional near-infrared spectroscopy to assess social information processing in poor urban Bangladeshi infants and toddlers. *Developmental Science, 22*(5), 1–15. https://doi.org/10.1111/desc.12839

Pereira, F., Mitchell, T., & Botvinick, M. (2009). Machine learning classifiers and fMRI: a tutorial overview. *NeuroImage, 45*(1 Suppl), 199-209. 19070668. doi: 10.1016/j.neuroimage.2008.11.007.

Quinn, P. C., Balas, B. J., & Pascalis, O. (2021). Reorganization in the representation of face-race categories from 6 to 9 months of age: Behavioral and computational evidence. *Vision Research, 179*(November 2020), 34–41. https://doi.org/10.1016/j.visres.2020.11.006

Rorden, C., Davis, B., George, M. S., Borckardt, J., & Fridriksson, J. (2008). Broca's area is crucial for visual discrimination of speech but not non-speech oral movements. *Brain Stimul, 1*(4), 383–385. https://doi.org/10.1901/jaba.2008.1-383.Broca

Ruxton, G. D., & Neuhäuser, M. (2013). Review of alternative approaches to calculation of a confidence interval for the odds ratio of a 2 × 2 contingency table. *Methods in Ecology and Evolution, 4*(1), 9–13. https://doi.org/10.1111/j.2041-210x.2012.00250.x

Scholkmann, F., Spichtig, S., Muehlemann, T., & Wolf, M. (2010). How to detect and reduce movement artifacts in near-infrared imaging using moving standard deviation and spline interpolation. *Physiological Measurement*, *31*(5), 649-662. 20308772. doi: 10.1088/0967-3334/31/5/004.

Sekiyama, K., Kanno, I., Miura, S., & Sugita, Y. (2003). Auditory-visual speech perception examined by fMRI and PET. *Neuroscience Research, 47*(3), 277–287. https://doi.org/10.1016/S0168-0102(03)00214-1

Tomalski, P., Ribeiro, H., Ballieux, H., Axelsson, E. L., Murphy, E., Moore, D. G., & Kushnerenko, E. (2013). Exploring early developmental changes in face scanning patterns during the perception of audiovisual mismatch of speech cues. *European Journal of Developmental Psychology, 10*(5), 611–624.

Tsurumi, S., Kanazawa, S., & Yamaguchi, M. K. (2019). Infant brain activity in response to yawning using functional near-infrared spectroscopy. *Scientific Reports, 9*(1), 1–9. https://doi.org/10.1038/s41598-019-47129-0

Venezia, J. H., Vaden, K. I., Rong, F., Maddox, D., Saberi, K., & Hickok, G. (2017). Auditory, Visual and Audiovisual Speech Processing Streams in Superior Temporal Sulcus. *Frontiers in Human Neuroscience, 11*(April), 1–17. https://doi.org/10.3389/fnhum.2017.00174

Weikum, W. M., Vouloumanos, A., Navarra, J., Soto-Faraco, S., Sebastián-Gallés, N., & Werker, J. F. (2007). Visual language discrimination in infancy. *Science, 316*(5828), 1159. https://doi.org/10.1126/science.1137686

Werker, J. F., & Tees, R. C. (1984). Cross-language speech perception: Evidence for perceptual reorganisation during the first year of life. *Infant Behavior and Development, 7*, 49–63. https://doi.org/10.1016/S0163-6383(02)00113-3

Xu, M., Hoshino, E., Yatabe, K., Matsuda, S., Sato, H., Maki, A., … Minagawa, Y. (2017). Prefrontal function engaging in external-focused attention in 5- to 6-month-old infants: A suggestion for default mode network. *Frontiers in Human Neuroscience, 10*, 676. https://doi.org/10.3389/fnhum.2016.00676

Zhao, H., Frijia, E. M., Rosas, E. V., Collins-Jones, L., Smith, G., Nixon-Hill, R., … Cooper, R. J. (2021). Design and validation of a mechanically flexible and ultra-lightweight high-density diffuse optical tomography system for functional neuroimaging of newborns. *Neurophotonics, 8*(01). https://doi.org/10.1117/1.nph.8.1.015011