1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25

Quantifying the individual auditory and visual brain response in 7- month-old infants watching a brief cartoon movie

Sarah Jessen¹ #, Lorenz Fiedler², Thomas F. Münte¹, & Jonas Obleser²

¹: Department of Neurology, University of Lübeck, Lübeck, Germany

²: Department of Psychology, University of Lübeck, Lübeck, Germany

Lead contact:

Dr Sarah Jessen, Department of Neurology, University of Lübeck, Ratzeburger Allee
160, 23562 Lübeck, Germany, Email: sarah.jessen@neuro.uni-luebeck.de
Phone: +49 451 3101 7449

Running title: Quantifying the individual infant brain response to real-life stimuli

Number of pages: 28

Number of figures: 7

Number of words (manuscript): 7573

Number of words (abstract): 245

Keywords: EEG; audiovisual; forward encoding models; temporal response function;
ecologically valid stimuli; developmental neuroscience

26 ABSTRACT

27 Electroencephalography (EEG) continues to be the most popular method to investigate
 28 cognitive brain mechanisms in young children and infants. Most infant studies rely on
 29 the well-established and easy-to-use event-related brain potential (ERP). As a severe
 30 disadvantage, ERP computation requires a large number of repetitions of items from
 31 the same stimulus-category, compromising both ERPs' reliability and their ecological
 32 validity in infant research. We here explore a way to investigate infant continuous EEG
 33 responses to an ongoing, engaging signal (i.e., "neural tracking") by using multivariate
 34 temporal response functions (mTRFs), an approach increasingly popular in adult-EEG
 35 research. $N=52$ infants watched a 5-min episode of an age-appropriate cartoon while
 36 the EEG signal was recorded. We estimated and validated forward encoding models
 37 of auditory-envelope and visual-motion features. We compared individual and group-
 38 based ('generic') models of the infant brain response to comparison data from $N=28$
 39 adults. The generic model yielded clearly defined response functions for both, the
 40 auditory and the motion regressor. Importantly, this response profile was present also
 41 on an individual level, albeit with lower precision of the estimate but above-chance
 42 predictive accuracy for the modelled individual brain responses. In sum, we
 43 demonstrate that mTRFs are a feasible way of analyzing continuous EEG responses in
 44 infants. We observe robust response estimates both across and within participants
 45 from only five minutes of recorded EEG signal. Our results open ways for
 46 incorporating more engaging and more ecologically valid stimulus materials when
 47 probing cognitive, perceptual, and affective processes in infants and young children.

48

49 INTRODUCTION

50 Neuroimaging studies in healthy human infants are subject to severe constraints, as
 51 participants cannot follow verbal instructions, show generally short attention spans,
 52 and overall tend to be not very cooperative. As functional magnetic resonance imaging
 53 (fMRI) studies are difficult to realize in infants (Ellis & Turk-Browne, 2018),
 54 electroencephalography (EEG) continues to be the most popular method to investigate
 55 cognitive brain mechanisms in very young children and infants.

56 To analyze the EEG signal, most studies in infants rely on the use of event-related brain
 57 potentials (ERPs). Accordingly, most infant EEG paradigms have been optimized for
 58 the computation of ERPs: This method necessitates that a few, carefully selected
 59 stimulus conditions are repeated multiple times to elicit and average a stereotypical
 60 brain response (i.e., an ERP) that can then be compared between conditions or between
 61 individuals. This leads to experimental designs that are often (a) highly unnatural and
 62 (b) have difficulties capturing the infants' attention for more than a few minutes.

63 However, in recent years and with the advent of modern computational capabilities,
 64 several new approaches to analyze EEG data have become available in adult EEG
 65 research. One such approach is the so-called "neural tracking", which seeks to
 66 compute and assess the relationship between the recorded EEG signal and an ongoing
 67 stimulus signal. The key ideas here are, first, naturally varying, non-repetitive stimuli,
 68 often movies (Bartels, Zeki, & Logothetis, 2008; Hasson, Nir, Levy, Fuhrmann, &
 69 Malach, 2004; Nishimoto et al., 2011) or naturally spoken conversation (Broderick,
 70 Anderson, Di Liberto, Crosse, & Lalor, 2018; Ding & Simon, 2013; Fiedler, Wöstmann,
 71 Herbst, & Obleser, 2019), which have higher ecological validity and arguably engage
 72 the participant qualitatively differently than artificial, isolated stimuli (Hamilton &
 73 Huth, 2018; Huk, Bonnen, & He, 2018; Matusz, Dikker, Huth, & Perrodin, 2018).
 74 Second, a mathematical framework (usually a variant of the general linear model) that
 75 allows to either "reconstruct" features of such a natural stimulus based on the ongoing
 76 brain response (so-called backward or decoding models), or to "predict" the measured
 77 ongoing brain response from features of the stimulus (so-called forward or encoding
 78 models; Dayan & Abbott, 2001; Naselaris, Kay, Nishimoto, & Gallant, 2011).

79 While the use of these advanced EEG analysis approaches has become rapidly
 80 mainstream in non-human and adult human neuroscientific research, it is still rare in
 81 infant research. This is unfortunate, since they not only have yielded important new
 82 insights in adult research and are likely to offer the same potential in infant studies,

83 but they may even provide higher gains in infancy research, which suffers from
84 notoriously low data quality and quantity. It may for instance reduce attrition rates, as
85 experimental designs can be optimized to be highly engaging for infant participants.
86 Rather than presenting hundreds of repetitions of very similar stimuli, which raises
87 the additional challenge of keeping a non-cooperative participant attending to the
88 screen, participants can be presented with constantly changing, engaging videos in
89 which stimuli are embedded.

90 Importantly, as in adult work, infant brain research has seen an increased interest in
91 the use of naturalistic settings over the past years. Recent research has for instance
92 demonstrated the feasibility of investigating interpersonal neural coupling in adult-
93 infant-interactions (Leong et al., 2017) or the use of oscillatory brain responses in
94 analyzing responses to dynamic social information (Jones, Venema, Lowy, Earl, &
95 Webb, 2015). While dynamic, naturalistic settings and experimental paradigms yield
96 important new insights into how brains behave and interact in real life rather than an
97 abstract laboratory setting, they inherently pose the additional challenge of hard-to-
98 predict and highly variant sensory input. Being able to directly relate a constantly
99 changing input to ongoing brain responses would therefore also be crucial for the
100 analysis of state-of-the-art ecologically valid experimental designs.

101 One particularly promising approach to do so is the use of multivariate temporal
102 response functions (mTRFs), which allow us to linearly link ongoing, continuous
103 environmental signals to simultaneously recorded brain responses in a
104 mathematically deterministic and computationally straightforward (i.e., non-iterative)
105 procedure. In adults, mTRFs have successfully been used to track the processing of
106 ongoing speech (e.g., Fiedler et al., 2019) as well as ongoing and naturalistic visual
107 input (O’Sullivan, Crosse, Di Liberto, & Lalor, 2017). Furthermore, Kalashnikova et al.
108 (2018) used mTRFs in infants to analyze the processing of ongoing auditory speech
109 signals, reporting a stronger cortical tracking for infant-directed compared to adult-
110 directed speech (Kalashnikova, Peter, Di Liberto, Lalor, & Burnham, 2018).

111 We here demonstrate the feasibility and utility of a forward encoding modelling
112 combined with non-repetitive complex multisensory stimulation in an infant
113 population. We presented 7-month-old infants with a 4’48’’ long age-appropriate
114 cartoon (one episode of the cartoon-show *Peppa Pig*) while recording the EEG. We
115 focused our analysis on the processing of three low-level physical stimulus
116 parameters; the auditory envelope, the motion content, and luminance. All three

117 parameters have been amply investigated in both infants and adults and are known to
118 elicit reliable ERP responses.

119 The auditory ERP response typically consists of a frontocentral P1-N1-P2-N2
120 sequence of responses, which can be clearly observed in adults and emerges in infancy
121 and early childhood (see e.g., Wunderlich & Cone-Wesson, 2006, for a review).
122 Compared to adults, infants tend to show a much less pronounced P1-N1 response,
123 and the overall response is dominated by a broad P2 response (Wunderlich, Cone-
124 Wesson, & Shepherd, 2006). The topographical distribution of the auditory evoked
125 potential in infants is comparable to that of adults and characterized by a broad
126 frontocentral distribution extending to parietal and temporal areas (Barnet, 1971).

127 The infant visual ERP to complex stimuli such as objects and faces comprises three
128 main components; the Pb, the Nc, and the Slow Wave (Webb, Long, & Nelson, 2005).
129 In particular, the Nc response, a frontocentral negativity typically observed between
130 400 and 800 ms after stimulus onset often linked to the allocation of attention has been
131 amply investigated (de Haan, Johnson, & Halit, 2003; Reynolds & Guy, 2012).

132 If we were successful in estimating auditory and visual brain responses using a
133 forward encoding model approach, we expect response functions comparable to
134 classical evoked brain responses. Furthermore, since the combined use of auditory and
135 visual regressors provides more information compared to the use of either regressor
136 alone, we expected a more consistent and reliable response function when using
137 auditory and visual regressors in one model.

138 Finally, while the predictive accuracy based on individual mTRFs computed on a
139 subset of the available data reveals consistencies *within* subjects (Fiedler et al., 2017;
140 Fiedler et al. 2019), a “generic” mTRF, that is, an average response function computed
141 across participants, reveals consistencies *across* subjects (O’Sullivan et al., 2015;
142 Mirkovic et al. 2015; Di Liberto & Lalor, 2017). Due to the limited amount of data
143 available in the infant cohort we aimed to explore the potential benefit from relying on
144 a generic model. Hence, we computed an averaged response function over $n-1$
145 participants and used this response function to model responses in the n th participant
146 (i.e. leave-one-out cross validation). We directly contrasted the predictive accuracy
147 obtained with these two approaches (individual mTRF versus generic mTRF) in the
148 present data set.

149 The present manuscript thus extends previous approaches on infant mTRFs
150 (Kalashnikova et al., 2018) by (a) using multisensory stimuli; (b) directly contrasting

151 infant and adult responses; (c) testing a large group of infants (n=52); and (d)
152 contrasting two different approaches to compute mTRFs, namely using a generic
153 response function compared to an individual response function.

154

155 METHODS

156 *Infant participants.* Fifty-two 7-month-old infants were included in the final sample
157 (age: 213 ± 8 days [mean $\pm$ standard deviation (SD)]; range: 200-225, 24 female). Not
158 untypical for infant studies (Stets, Stahl, & Reid, 2012), an additional 39 infants had
159 been tested but could not be included in the final sample. Note also that directly prior
160 to the experiment reported here, infants had already participated in a 5-10-minute-
161 long ERP experiment on visual emotion perception (see below), further contributing
162 to the drop-out rate since infants often became fussy or tired after the first experiment.
163 In detail, infants were excluded because they did not watch the complete video (n=24);
164 were too fussy to watch the video at all (n=10); did not contribute at least 100 s of
165 artifact-free data (n=3); had potential neurological problems (n=1); or because of
166 technical problems during the recording (n=1).

167 All infants were recruited via the maternity ward at the local hospital
168 (Universitätsklinikum Schleswig-Holstein); were born full-term (38-42 weeks
169 gestational age); had a birth weight of at least 2500 g; had no known neurological
170 deficits; and grew up in predominantly German-speaking households. The study was
171 conducted according to the declaration of Helsinki, approved by the ethics committee
172 at the University of Lübeck, and parents provided written informed consent.

173 *Adult reference sample.* In addition, we collected data from a reference sample of n = 33
174 adult participants. Data from n=5 were excluded due to technical difficulties during
175 the recording (n=2) or failure to contribute at least 100 s of artifact-free data (n=3),
176 leading to a final sample of n=28 (mean age: 50 years; range: 21-69, 16 female).

177 *Stimulus.* As stimulus material we used one episode of the cartoon show Peppa Pig
178 (Episode "Peppa Pig - The new car", dubbed in German), an age-appropriate cartoon
179 featuring a family of pigs and their daily life. Duration of the entire movie clip was
180 4'48'', that is, 269 s or 6451 frames. Sound and visual parameters were not manipulated
181 in any way.

182 *Procedure-Infants.* After arrival in the laboratory, parents and infant were familiarized
183 with the environment and parents were informed about the study and signed a

184 consent form. The EEG recording was prepared while the infant was sitting on his/her
185 parent's lap. For recording, we used an elastic cap (BrainCap, Easycap GmbH) in
186 which 27 Ag/AgCl-electrodes were mounted according to the international 10-20-
187 system. An additional electrode was attached below the infant's right eye to record the
188 electrooculogram, which however we did not use for the present analyses. The EEG
189 signal was recorded with a sampling rate of 250 Hz using a BrainAmp amplifier and
190 the BrainVision Recorder software (both Brain Products).

191 For the EEG recording, the infant was sitting in an age appropriate car seat (Maxi Cosi
192 Pebble) positioned on the floor. As part of a larger study, a t-shirt was positioned over
193 the chest area of the infants. The t-shirt had either previously been worn by the infant's
194 mother (n=19) or by the mother of a different same-aged infant (n=14) or had not been
195 worn before (n=19). This modulation was not of main interest to the present study and
196 will not be analyzed or reported here in further detail.

197 In front of the infant (approximately 60 cm from the infant's feet), a 24-inch monitor
198 with a refresh rate of 60 Hz was positioned at a height of about 40 cm (bottom edge of
199 the screen), resulting in a horizontal visual angle of approximately 27.5° and a vertical
200 visual angle of approximately 17.6°. Left and right of the monitor, loudspeakers
201 (Logitech X-140) were positioned and set to a comfortable level of loudness. When the
202 infant was attending to the screen, the video was started and played without
203 interruption until the end of the episode. The parent was seated approximately 1.5 m
204 behind the infant and was instructed not to interact with the infant during the video.
205 In case the infant became too fussy and started crying during the video, the video was
206 aborted and the infant was excluded from further analysis.

207 Before this video presentation, infants had been presented with a series of photographs
208 displaying happy and fearful facial expressions as part of the larger, maternal-odor
209 study. Again, the results of this part of the study will not be further analyzed here.

210 *Procedure-Adults.* Adult participants were presented with the same "Peppa Pig" movie
211 after they had already participated in one of several unrelated EEG studies. They were
212 informed about the study and signed a consent form. For recording the EEG signal,
213 we used 64 Ag/AgCl active scalp electrodes positioned in an elastic cap according to
214 the international 10-20-system. The EEG signal was recorded with a sampling rate of
215 1000 Hz using an ActiChamp amplifier and the BrainVision Recorder software (Brain
216 Products).

217 Adult participants sat in a soundproof and electrically shielded chamber (Desone) in
218 a comfortable chair approximately 1 m away from a 24-inch monitor with a refresh
219 rate of 60 Hz on which the video was presented. Sound was presented from the same
220 loudspeaker models used in the infant study, also positioned left and right to the
221 screen (Logitech X-140).

222 *Analysis.* Unless noted otherwise, the analysis protocol including the preprocessing
223 was identical for infant and adult data. We analyzed the data using Matlab 2013b (The
224 MathWorks, Inc., Natick, MA), the Matlab toolbox FieldTrip (Oostenveld, Fries, Maris,
225 & Schoffelen, 2011), and the multivariate temporal response function (MTRF) toolbox
226 (Crosse, Di Liberto, Bednar, & Lalor, 2016).

227 *Preprocessing.* The data were referenced to the average of all electrodes (mean
228 reference), filtered using a 100-Hz-lowpass and a 1-Hz-highpass filter, and segmented
229 into 1-sec-epochs. To detect epochs contaminated by artifacts, the standard deviation
230 was computed in a sliding window of 200 msec. If the standard deviation exceeds 80
231 mV at any electrode, the entire epoch was discarded, and if less than 100 artifact-free
232 epochs remained, the participant was excluded from further analysis. An independent
233 component analysis (ICA) was computed on the remaining concatenated data.
234 Components were inspected visually by a trained coder (S.J.) and rejected if classified
235 as artefactual (infants: 5 ± 2 components per participant [mean $\pm$ SD], range 1–10;
236 adults: 26 ± 5 , range 11–36). A 1–10 Hz bandpass filter was applied to the cleaned data.
237 Adult data were downsampled to the infant-data sampling frequency of 250 Hz.

238
239 *Extraction of stimulus regressors.* Regressors characterizing motion, luminance,
240 and the sound envelope were extracted from the stimulus video. Exemplary excerpts
241 of audio, luminance, and motion regressors are shown in Fig 1B.

242 To compute a regressor of average luminance across all pixels, the weighted
243 sum of the rgb values for each frame was computed using Matlab (Bartels et al., 2008).

244 To compute a regressor of average motion across all pixels, each video frame
245 was converted to grey-scale, and the difference between two consecutive frames was
246 computed. Then, the mean across all pixels for which this difference was larger than
247 10 (to account for random noise, see e.g. Jessen & Kotz, 2011; Pichon, de Gelder, &
248 Grèzes, 2009) was computed.

249 To compute a regressor of sound envelope, the audio soundtrack of the video
250 was extracted and submitted to the NSL toolbox, an established preprocessing pipeline
251 emulating important stages of auditory peripheral and subcortical processing (Ru,

252 2001). The output of this toolbox resulted in a representation containing band-specific
 253 envelopes of 128 frequency bands of uniform width on the logarithmic scale with
 254 center frequencies logarithmically spaced between 0.1 and 4 kHz. To obtain the
 255 broadband temporal envelope of the audio soundtrack, these band-specific envelopes
 256 were then summed up across all frequencies to obtain one temporal envelope.
 257 Following earlier own and others' approaches, we used the first derivative of the half-
 258 wave rectified envelope as the final audio regressor (for details see Fiedler et al., 2017).
 259 The result is a pulse-train-like series of peaks where, across frequency bands, the
 260 acoustic energy rises most steeply, reflecting "acoustic edges" such as syllable onsets.

261

262 *Stimulus parameters.* As expected for a child-friendly cartoon movie, frame-to-frame
 263 fluctuations in luminance were small. On average, the change in luminance from one
 264 frame to the next was 0.35 units per frame (range 0–53, median = 0.05). Note that this
 265 deviates from previous studies where the entire dynamic range of luminance (i.e.,
 266 black to white) was used to quantify the temporal response function in the adult EEG
 267 response (e.g., Lalor, Pearlmutter, Reilly, McDarby, & Foxe, 2006; Vanrullen &
 268 MacDonald, 2012) or in non-human animal electrophysiological responses (Ringach &
 269 Shapley, 2004). In contrast, the luminance-derived motion regressor yielded sizable
 270 variance, with a mean frame-to-frame change of 38 units (range 0–192, median = 36).

271 In sum, while variance in the luminance regressor was small, both motion and
 272 audio regressors showed considerable and promising degrees of variance.

273 Lastly, all regressors were downsampled (audio) and interpolated (motion,
 274 luminance), respectively, to the EEG sampling frequency of 250 Hz. In all regressors,
 275 time periods in which no EEG data was available as a result of artefact rejection during
 276 preprocessing were zero-replaced. Finally, EEG data and physical regressors were
 277 aligned and available for the linear model analysis.

278

Figure 1. Physical properties of stimulus regressors. A) shows four exemplary stills from the movie used as stimulus material. B) shows an example of a 5-s-long stretch of the audio (orange), motion (purple), and luminance regressors (blue). Below, the frequency spectrum of the stimulus regressors is depicted; while frequencies < 10 Hz appear to be dominant in the audio regressors, no such dominance can be observed for the other regressors. C) shows an overview of the analysis approach. During training, stimulus regressors and the EEG signal of $n-1$ participants was used to compute a generic response function (left part). During testing, this generic response function was used to predict the EEG response of the n th participant, which was then compared to the actual EEG response of that participant. See main text for further details.

Temporal response functions (TRF). To quantify the degree to which the EEG of 7-month-olds (as well as adults) can be expressed as a linear response to stimulus features, we used regularized regression (with ridge parameter λ) as implemented in the mTRF toolbox (Crosse et al., 2016). The key idea here is to estimate a temporal response function (TRF), that is, a set of time-lagged weights g , with which a regressor s (here, the physical stimulus features) would need to be convolved (i.e. multiplied and summed) in order to optimally predict the measured EEG response r .

More specifically, we used a forward encoding model approach. In a first pass, we aimed to maximize the predictive accuracy of such a model by estimating so-called “generic” models, that is, we predicted the EEG data of an n th participant based on a “generic” temporal response function (TRF) from $n-1$ participants to the auditory or visual stimulus signal. Since changes in the EEG signal are not likely to occur simultaneously with changes in the stimulus signal but rather with an (unknown) time

lag, predictions were computed over a range of time lags between 200 ms earlier than the stimulus signal and 1000 ms later than the stimulus signal.

Choosing the optimal regularization parameter λ . To obtain the optimal regularization parameter λ for each stimulus regressor separately, as well motion and audio simultaneously, we trained the respective model on the EEG data of each participant using a variety of λ values between 10^{-5} and 10^5 . We increased the exponent in steps of 0.5, and used the resulting models to predict the EEG signal for each participant. Hence, we obtained a total of twenty different models (and predictions) based on the different λ parameters. For each of these models, we then computed the mean response function across $n-1$ participants and used this response function to predict EEG response of the n th participant (i.e., n -fold leave-one-out crossvalidation). Finally, we computed the predictive accuracy (i.e., Pearson's correlation coefficient r between the predicted EEG response and the actual EEG response) for each participant, resulting in one accuracy value for each electrode (27 for infants, 64 for adults) per participant and stimulus parameter for each λ value. For each participant, stimulus parameter, and electrode, we selected the λ value maximizing *predictive* accuracy, that is, the value for which the model yielded the highest correlation between the predicted and the actual EEG. Finally, we computed the mean regularization parameter λ value by averaging across all electrodes and participants (see Table S1). This procedure was done separately for infant and adult participants, resulting in different λ values for infants and adults.

These optimal λ parameters were used in the following to train the model, resulting in separate response functions for each stimulus parameter. For each of the three physical stimulus parameters (luminance, motion, audio) we computed a separate model. In addition, we computed a model using both, motion and audio, as regressors ("joint audio-motion model"). We chose not to include luminance in this model, as the regressor for luminance did not yield any reliable model in itself (see results).

Evaluation of temporal response functions. For statistical evaluation of the resulting response functions, we computed a cluster-based permutation test with 1000 randomizations, testing the obtained response functions against zero. A cluster was defined along the dimensions time and electrode position, with the constraint that a cluster had to extend over at least two adjacent electrodes. A type-1-error probability of less than .05 was ensured at the cluster level (Maris & Oostenveld, 2007). Note that the number and extent of the largest clusters in the originally observed data can be

339 compared to clusters as obtained from random permutations of the data. This
340 constitutes the actual test at the cluster level and it protects the family-wise error at the
341 desired type-I-error rate, here also 5 %.

342 In addition, to assess internal validity of our model predictions on an
343 individual basis, we computed three different predictive accuracies per participant.
344 First, for each participant n , we computed the correlation between the predicted
345 response generated on a model trained on $n-1$ participants and the actual EEG
346 response of n ("generic model").

347 Second, rather than relying on the generic model based on $n-1$ participants, we
348 computed an individual response function for each participant ("individual model").
349 To that end, 80 % of the available data for a given participant were used to train the
350 model, and the resulting response function was then correlated with the response
351 observed in the remaining 20 % of the data.

352 Third, a *permuted* or null predictive accuracy ("shifted control") was obtained.
353 Before calculating accuracy this way, we shifted the actual EEG response for
354 participant n in steps of 2 s (in order to ensure to exceed the potential autoregressive
355 structure of the EEG data) and computed the correlation between the shifted EEG
356 signal and the predicted response, based on the generic model trained on $n-1$
357 participants.

358 To further assess the temporal unfolding of the neural tracking, changes in
359 predictive accuracy across time were assessed using a sliding window over a range of
360 timelags from -200 to 1000 ms (see also Fiedler et al., 2019). In particular, we used a
361 sliding window with a width of 48 ms (12 samples) and an overlap of 24 ms (6 samples)
362 and computed the predicted EEG signal and its correlation with the actual recorded
363 EEG signal for each window. Finally, we selected the time-window of maximal
364 predictive accuracy, computed the predicted EEG signal for this optimized time-
365 window, and directly compared the obtained results to the results obtained by the
366 generic model based on the entire range of time lags.

367 Finally, we used t-tests to directly compare infant and adult response functions
368 at each sampling point over a range of time lags between -200 ms and 1000 ms,
369 correcting the results for multiple comparison by using the false-discovery rate
370 procedure as suggested by Benjamini and Hochberg (1995).

371

372 RESULTS

373 *Temporal response function.* We computed a generic temporal response function for each
374 stimulus regressor as well as the audio and motion regressor combined (joint audio-
375 motion model).

376 We observed a clearly defined response function using the audio regressor and
377 the motion regressor (Figure 2 and Figure S2/S3 for adults), while no clear response
378 function could be obtained using the luminance regressor for either infants or adults
379 (Figure S1). While Figure 2 shows the respective response functions obtained from a
380 model which included both regressors (joint audio-motion model), comparable
381 response functions resulted when using either of the regressors in isolation. When
382 directly comparing the predictive accuracy obtained by the joint audio-motion model
383 to that of the models using either regressor in isolation, we observed a higher
384 predictive accuracy for the joint model for both, the audio and the motion regressor,
385 in infant (joint model vs. auditory only: $t(51) = -4.4$, $p < .001$; joint model vs. motion
386 only: $t(51) = -7.77$, $p < .001$). In the adult data, this was only the case for the contrast
387 between the joint model and the motion-only model (joint model vs. auditory only:
388 $t(27) = -.13$, $p = .20$; joint model vs. motion only: $t(27) = -8.44$, $p < .001$).

389 Interestingly, while a clearly defined response was visible for both, the audio
390 and motion regressor, the amplitude of the response function for the motion regressor
391 was much smaller compared to the amplitude of the audio response function.

392 *Cluster-based permutation test.* We computed a cluster-based permutation test
393 comparing the temporal response function obtained using the motion, luminance, and
394 audio regressor as well as the motion and audio regressor simultaneously. We did not
395 observe any significant cluster using the luminance regressor for either infants or
396 adults. In contrast, we did obtain multiple significant clusters, indicating a positive or
397 negative deviation from zero, for the motion and audio regressor, both when included
398 separately as well as in combination (see supplementary material for a full list of the
399 results of the cluster-based permutation test using audio and motion regressor
400 separately as well as in combination for infants and adults, Figure 2B for infant results
401 and S2B/ S3B for adult results). The resulting clusters confirm the deflections observed
402 in the auditory and motion response function (Figure 2A).

403

404 *Comparing the infant and adult brain responses.* When comparing infant and adult
 405 response functions (Figure 3), similarities as well as striking differences emerge.
 406 Overall, amplitudes of the response functions are comparable for infants and adults,
 407 both showing the already mentioned larger amplitudes for audio regressors and
 408 smaller amplitudes for motion regressors. For both, infants and adults, the auditory
 409 response function is marked by a prominent frontocentral positivity (250–500 ms for
 410 infants, 300–450 ms for adults). While this response appears to be slightly longer for
 411 infants, overall, both latency and topography indicate a comparable response for
 412 infants and adults. In contrast, the infant auditory response function lacks a second,
 413 earlier and more central positivity, which can be observed between 150 and 250 ms in
 414 the adult auditory response function.

415 For the motion response function, both infants and adults show two frontal /
 416 frontocentral positivities (250– 450 and 700–900 ms for infants and 250–350 and 450–
 417 550 ms for adults). Hence, infants and adults show a comparable response, though the
 418 infant response appears to be much slower and less temporally modulated.

419 When statistically comparing infant and adult responses directly, significant
 420 differences can be observed at a time lag of around 200 ms and 300 ms in the auditory
 421 response function (Figure 3A). In contrast, the motion response function differs
 422 between infants and adults at multiple lags (Figure 3B).

423

424

425 *Figure 2. Response function (using motion and audio regressor simultaneously) for infant*
426 *participants. The left panel shows the auditory response function, while the right panel*
427 *shows the motion response function. A) depicts the mean mTRF (mean $\pm$ SEM)*
428 *computed across all participants, averaged over FC5, FC6, F3, and F4 (auditory)*
429 *respectively F3, Fz, F4, C3, Cz, and C4 (motion) with the included electrodes marked*
430 *by black dots. Electrodes were chosen based on visual inspection of the topography*
431 *and the cluster test (B). Topographic representations for are shown 0–250 ms, 250–500*
432 *ms, and 500–700 ms (auditory) and 50–250 ms, 250–450 ms, 450–700 ms, and 700–900*
433 *ms (motion). B) displays the results of the cluster-based permutation test, comparing*
434 *the response function shown in A) to zero. Positive deviations are displayed in red,*
435 *while negative deviations are shown in blue. In the bottom part of B), the same clusters*
436 *as in the top part of B) are shown as topographic distributions, along with the summed*
437 *t-value across the cluster.*

Statistical comparison between infant and adult response functions

Figure 3. Comparison of infant and adult response functions. Mean mTRF for infant (in blue) and adult (in red) participants are shown for the audio regressor (A) and the motion regressor (B). The infant response functions and topographical representations are identical to those shown in Fig 2A and 3A for audio and motion regressors, respectively. Responses are averaged across the same electrodes for adults and infants, namely FC5, FC6, F3, and F4 for A) and F3, Fz, F4, C3, Cz, and C4 for B). The topographic representations of adult responses correspond to those in the supplementary material, namely 50–150 ms, 150–250 ms, 250–300 ms, and 300–450 ms for A) and 50–250 ms, 250–350 ms, 350–450 ms, 450–550 ms, and 550–800 ms for B). The graph in the bottom part depicts t-values of the mean infant – adult difference. Periods in which t-tests resulted in a p value <.05 are marked by green asterisks, periods in which these tests survived correction for false discovery rate are marked by orange asterisks.

Temporal unfolding of predictive accuracy. To further investigate the temporal unfolding of the response functions, we computed the predictive accuracy (i.e., the correlation between the predicted and the actual EEG response) using a sliding window of 48 ms (Figure 4). Both, infants and adults, show the highest predictive accuracy at time lags between 200 and 400 ms (“optimized time windows”). However, when using only these optimized time windows to compute the predictive accuracy, a different pattern between infants and adults emerges. In adults, predictions based on this optimized time window yielded a higher predictive accuracy compared to the generic model encompassing the entire range of time lags ($t(27) = -6.51, p < .001$). In infants, however, predictions based on this optimized time window yielded a lower predictive accuracy compared to the generic model encompassing the entire range of time lags ($t(51) = 7.39, p < .001$).

Figure 4. *Temporal unfolding of predictive accuracy.* Depicted in gray is the correlation between predicted and actual EEG in sliding time-windows of 48 ms (with 24 ms overlap, left panel: infants, right panel: adults; mean $\pm$ SEM). The area marked in orange denotes the time range used as optimized time window. Predictive accuracy was computed over the optimized time window and is shown by orange circles (next to the predictive accuracy yielded by the entire time window in blue for comparison; mean accuracies with 95 % confidence intervals are shown in black).

Generic vs. individual response functions. The results discussed above rely on a generic model, that is, an average TRF computed based on data from $n-1$ participants was convolved with the n th participant's data in order to obtain an predicted EEG trace for this subject (see Di Liberto & Lalor, 2017). An alternative approach (and in fact preferable, if enough data for per subject is available; e.g., Fiedler et al., 2019; O'Sullivan et al., 2017) computes an individual model based on a subset of an individual's data and compare the resulting predictions to the remaining data.

As expected, individual models showed a larger variance compared to the generic model (Figure 5-7; see S4 and S5 for data from adult participants), but both, generic model and individual model result in correlations clearly above zero (with the exception of luminance, where no reliable prediction was possible for either mode, see Figure 5C; results of the t-tests against zero for the joint-model depicted in Figure 7: generic model_{infant}: $t(51) = 15.72$, $p < .001$; generic model_{adult}: $t(27) = 11.49$, $p < .001$; individual model_{infant}: $t(51) = 5.60$, $p < .001$; individual model_{adult}: $t(27) = 8.33$, $p < .001$).

When both, audio and motion regressor were included (Figure 6), the generic model resulted in a higher correlation compared to the individual model for infant participants ($t(51)=3.76$, $p < .001$); 37 participants showed a higher correlation with the generic model while only 15 participants showed a higher correlation with the individual model (Figure 6, right panel). Nevertheless, both approaches provide converging results, as suggested by a significant positive correlation between the

predictions obtained the generic model and those obtained with the individual model ($r=.36, p=.009$; infants, joint regressor). When using only the motion regressor (Figure 5B), the correlations were also higher for the generic compared to the individual model ($t(51)=3.50, p<.001$), while for the auditory regressor (Figure 5A), this difference was less pronounced ($t(51)=1.82, p=.07$).

As a control analysis, a generic model using temporally shifted (i.e., purposefully misaligned) versions of the actual EEG signal (1,000 iterations) did yield substantially lower predictive accuracy values.

Finally, we directly compared the use of the generic compared to the individual model between infants and adults (Figure 7). For both, the generic and the individual model, the adult group showed a higher predictive accuracy (generic model: $t(78) = -2.46, p=.016$; individual model: $t(78) = -3.54, p<.001$).

Figure 5. Predictive Accuracy (r) between model and EEG response for infant participants. The recorded individual EEG response was correlated with three different parameters using Pearson's correlation coefficient for the audio regressor (A), motion regressor (B), and luminance regressor (C). On the left, the correlation between the recorded EEG responses of participant n and the response predicted by the generic model based on the remaining $n-1$ participants is shown for each participant. In the middle, the correlation between the model trained on the first 80 % of the data available for each participant and used to predict the remaining 20 % from that participant and the actual EEG response recoded from that participant is shown. The right column shows the correlation between the prediction generated by the generic model and the recorded EEG data shifted in a circular way in steps of 2 s as a control condition (averaged over all possible shifts). Correlations are shown for each infant participant (in colors) as well as the mean correlation with 95% CI (confidence interval) across all participants (in black).

543

544

545 *Figure 7. Predictive accuracy for infants and adults in a joint audio-motion model. A) shows*
546 *the individual correlations using Pearson's correlation coefficient for infants (blue) and*
547 *adults (orange) using the generic model. B) shows the individual correlations using*
548 *Pearson's correlation coefficient for infants (blue) and adults (orange) using the*
549 *individual model. Mean accuracies with 95 % confidence intervals are shown in black.*

550

551 DISCUSSION

552 We here investigated the use of a variant of forward encoding models (multivariate
553 temporal response functions, mTRFs) to analyze infant brain responses to a continuous
554 complex audiovisual stimulus, namely a 5-minute cartoon movie. We observed clearly
555 defined response patterns to both the auditory as well as the motion content, but no
556 predictive response function for changes in luminance was found.

557 Our results demonstrate that the simultaneous acquisition of individual brain
558 responses to different sensory modalities is possible in the infant brain, opening new
559 avenues for ecologically valid multisensory research paradigms in developmental
560 neuroscience. Furthermore, our results suggest that a generic model derived from a
561 larger set of unrelated infant data is as good or slightly better compared to an
562 individual model in predicting the individual brain response, especially in cases where
563 only limited data is available. This points to the further utility of such an approach in
564 developmental and at-risk populations.

565 *Motion and Audio.* For both, motion and audio information in the cartoon movie, we
566 found a clearly defined response in both infants and adults. The observed responses
567 are largely consistent with patterns typically reported in more traditional event-
568 related brain potentials. In particular, the frontocentral negativity between 450 and 700
569 ms observed in the infant brain responses linked to the motion regressor corresponds
570 in timing, shape, and topography to the Nc component, an infant ERP component that

571 can routinely be observed in visual paradigms and has been linked to attention
572 allocation (Webb et al., 2005). Furthermore, the bifocal frontal positivity observed in
573 the infants' brain response linked to the auditory envelope shows a strong similarity
574 to the commonly reported P2 response in infant auditory brain responses (Wunderlich
575 et al., 2006), though the peak can be observed around 400 ms and therefore somewhat
576 later than a typical P2 response.

577 Similar comparisons can be drawn for observed response functions in the adult group.
578 The auditory response function is characterized by an P1-N1-P2-N2 pattern, which
579 resembles the typical adult auditory onset response (Picton, 2013). As in the infants,
580 the response appears to be slower though, with the first positivity peaking around 200
581 and the second around 350 ms. The topographical distribution, namely a broad
582 frontocentral activation, is consistent with typical auditory ERP responses (Picton,
583 2013). Importantly, the auditory response function also resembles the pattern that has
584 been reported in prior studies investigating neural tracking of attended auditory
585 information in adults (e.g. Fiedler et al., 2019).

586 The adult motion response function shows a pattern characterized by two positivities,
587 one peaking around 300 ms, the other around 500 ms. The topographical distribution
588 is again primarily frontocentral, though a positivity around 400 ms can be seen at
589 posterior and occipital electrodes (see Figure 3 and Supplementary Figure S3). Prior
590 comparable studies investigating the processing of continuous naturalistic stimuli
591 have also reported a response profile characterized by two positivities, although at
592 shorter latencies and with primarily occipital distributions (see e.g. O'Sullivan et al.,
593 2017).

594 The direct comparison of infant and adult brain responses (Figure 3) may
595 provide insight into developmental changes. In response to the auditory envelope,
596 both infants and adults show a prominent frontal negativity peaking around 400 ms.
597 Notably, however, the adults show an additional central positivity around 200 ms,
598 which is missing in the infant response. This corroborates and replicates known
599 developmental changes commonly observed in auditory evoked responses when
600 comparing infants and adults (Wunderlich & Cone-Wesson, 2006). Considering the
601 motion response, the correspondence between infant and adult response is less
602 straight-forward. While the adult response is characterized by two frontocentral
603 positivities, one peaking around 300 ms and the other around 500 ms, the infant
604 response is dominated by one frontocentral peak around 400 ms.

605 Importantly, we used both, generic response functions as well as individual
606 response functions to predict the EEG signal. When using both, the motion and the
607 auditory regressor, performance was significantly better for the generic compared to
608 the individual model. When using only the auditory regressor, the same pattern was
609 visible but the difference only marginally significant. Note, however, that both, generic
610 and individual models generated predictions that were significantly above chance
611 level. This demonstrates two important things. First, five minutes of EEG recording
612 are sufficient to compute reliable models, both on an individual level as well as across
613 participants as a generic model. This is not only true for EEG data obtained from
614 healthy adults but also for data obtained from populations providing notoriously
615 noisy signal, such as infants. Second, brain responses across participants, both infants
616 and adults, are sufficiently similar to generate a model that can successfully predict a
617 new infant's brain response, yielding even better outcomes compared to the individual
618 model.

619 *Limitations and future studies.* The present study provides an important step and proof
620 of feasibility for using mTRFs to analyze infant EEG data in response to complex and
621 dynamic audiovisual stimulus material. This offers a whole host of new possibilities
622 in the investigation of infant's brain responses in their natural environment.

623 One important advantage of using mTRFs compared to classical ERPs
624 approaches is the possibility of using more complex and naturalistic experimental
625 settings. These might include the processing of continuous auditory signals (as
626 demonstrated by Kalashnikova et al., 2018), or audiovisual video material, as in the
627 present study. In contrast to ERPs, which require the repetitive presentation of short
628 stimuli, with mTRFs, it is possible to investigate the brain's entrainment over a longer
629 period of time and in a much closer approximation of natural settings. Taking this last
630 aspect a step further, mTRFs thereby also offer a potential approach to analyze brain
631 responses in live interactions in which the live input the infant receives (or also the
632 infant's responses) are recorded and used as a regressor in the subsequent analysis.
633 Such an approach would provide an important tool in investigating the neural bases
634 of social interactions, a topic that has received increasing interest in recent years also
635 in developmental neuroscience (see e.g., Leong et al., 2017; Wass et al., 2018).

636 From a research-design point of view, the use of continuous stimulation (in
637 contrast to short repetitive stimuli needed for ERPs) also allows for the design of more
638 engaging experiments. This is particularly important for populations characterized by
639 short attention spans and overall low compliance, such as young children, but also

640 patient groups. In addition, such participant groups are often also characterized by a
641 low signal-to-noise ratio in the EEG signal and only limited data availability; here, we
642 demonstrate that 5 minutes of recording are sufficient to compute reliable responses,
643 highlighting a further practical advantage of using mTRFs to analyze EEG data in
644 special populations.

645 While the present study provides an important lead, it also raises several new
646 questions. One important feature of the present study is that we used the
647 unmanipulated cartoon video material. While this makes for an ecologically valid and
648 easy-to-obtain stimulus, it comes with the caveat of a lack of control for stimulus
649 properties.

650 Notably, while we did observe a clear-cut response to the motion and the auditory
651 regressor, we did not find a reliable response to the changes in luminance. The most
652 likely explanation for this discrepancy is the lack in variance in the luminance content.
653 While the motion and the auditory regressor showed large-amplitude changes
654 throughout the video (e.g., average motion change between frames = 38 units), average
655 luminance of this cartoon movie remained fairly constant (average luminance change
656 between frames = 0.35 units). Previous studies targeting neural responses to
657 luminance change (in adults) typically used considerably more pronounced black-
658 white contrast (Lalor et al., 2006; Vanrullen & MacDonald, 2012). Hence, the luminance
659 changes in the stimulus material were likely too small to elicit any robust change in
660 brain response. Future studies explicitly varying the luminance content are therefore
661 necessary to investigate the applicability of mTRFs to other visual stimulus parameters
662 in infants.

663 Also, we operationalized motion as change in pixel from one frame to the next.
664 This means that the motion regressor not only reflected the actual motion of the objects
665 and persons depicted in the video but also cuts in the video. For the present purpose,
666 we did not differentiate between these two possibilities of motion. Furthermore, in the
667 present set-up, we did not control for degree to which individual infants constantly
668 attended visually to the screen; future studies combining EEG recordings with eye
669 tracking might therefore further improve the predictive accuracy of the visual models.

670 Building upon the present results, a next step would therefore be to
671 purposefully manipulate such parameters. By using stimulus material designed to
672 encompass a larger variance in luminance and/or no cuts in the video, it should for
673 instance be possible to observe brain response to changes in luminance and motion
674 responses that can be clearly linked to actual motion rather than video cuts. Such an

675 approach could for instance provide valuable new insights into the processing of
676 biological motion (Marshall & Shipley, 2009; Reid, Hoehl, & Striano, 2006).

677 Furthermore, in the present study, we did not contrast different conditions,
678 neither within infant nor between different groups of infants. Having demonstrated
679 the feasibility of using encoding models to model brain responses for this type of
680 complex audiovisual stimuli, the next step would certainly be to utilize this approach
681 to investigate differences in processing between (a) different types of stimulation or
682 (b) different groups of infants.

683 *Conclusion.* The present data demonstrate that forward encoding models based on the
684 multivariate temporal response function (mTRF) pose a valuable and versatile tool in
685 quantifying and disentangling complex audiovisual brain responses and the according
686 perceptual processes in infancy. Our results open way for applications to a variety of
687 research areas not only in early development, but also in other special populations
688 characterized by short attention spans and low cooperativeness, including research in
689 severely impaired neurological patients. New paradigms could not only entail
690 complex multisensory perception, but extend to dynamic social interactions. As such,
691 mTRF approaches to infant data analysis will allow developmental researchers to
692 devise more engaging and thereby more easily applicable experimental set-ups for
693 infancy research.

694

695 ACKNOWLEDGEMENTS

696 We thank all the families for participating, Leonie Emmerich, Aylin Ulubas, Franziska
697 Scharata, and Anne Hermann for help with the data acquisition, and the German
698 Research Foundation (DFG) for funding to SJ (JE 781/1-1 & 2). Three reviewers
699 provided valuable and constructive feedback.

700 DATA ACCESSIBILITY.

701 Data will be made available upon publication on Open Science Framework (OSF).

702

703 BIBLIOGRAPHY

- 704 Barnett, A. B. (1971). Eeg audiometry in children under three years of age. *Acta Oto-*
705 *Laryngologica*. <https://doi.org/10.3109/00016487109122450>
- 706 Bartels, A., Zeki, S., & Logothetis, N. K. (2008). Natural vision reveals regional
707 specialization to local motion and to contrast-invariant, global flow in the human
708 brain. *Cerebral Cortex*. <https://doi.org/10.1093/cercor/bhm107>
- 709 Benjamini, Y., & Hochberg, Y. (1995). Controlling the False Discovery Rate: A Practical
710 and Powerful Approach to Multiple Testing. *Journal of the Royal Statistical Society:*
711 *Series B (Methodological)*. <https://doi.org/10.1111/j.2517-6161.1995.tb02031.x>
- 712 Broderick, M. P., Anderson, A. J., Di Liberto, G. M., Crosse, M. J., & Lalor, E. C. (2018).
713 Electrophysiological Correlates of Semantic Dissimilarity Reflect the
714 Comprehension of Natural, Narrative Speech. *Current Biology*.
715 <https://doi.org/10.1016/j.cub.2018.01.080>
- 716 Crosse, M. J., Di Liberto, G. M., Bednar, A., & Lalor, E. C. (2016). The Multivariate
717 Temporal Response Function (mTRF) Toolbox: A MATLAB Toolbox for Relating
718 Neural Signals to Continuous Stimuli. *Frontiers in Human Neuroscience*.
719 <https://doi.org/10.3389/fnhum.2016.00604>
- 720 Dayan, P., & Abbott, L. (2001). *Theoretical Neuroscience: Computational and Mathematical*
721 *Modeling of Neural Systems*. Cambridge, MA: MIT Press.
- 722 de Haan, M., Johnson, M. H., & Halit, H. (2003). Development of face-sensitive event-
723 related potentials during infancy: a review. *International Journal of Psychophysiology*,
724 51(1), 45–58. Retrieved from <http://www.ncbi.nlm.nih.gov/pubmed/14629922>
- 725 Di Liberto, G. M., & Lalor, E. C. (2017). Indexing cortical entrainment to natural speech at
726 the phonemic level: Methodological considerations for applied research. *Hearing*
727 *Research*. <https://doi.org/10.1016/j.heares.2017.02.015>
- 728 Ding, N., & Simon, J. Z. (2013). Adaptive Temporal Encoding Leads to a Background-
729 Insensitive Cortical Representation of Speech. *Journal of Neuroscience*.
730 <https://doi.org/10.1523/jneurosci.5297-12.2013>
- 731 Ellis, C. T., & Turk-Browne, N. B. (2018). Infant fMRI: A Model System for Cognitive
732 Neuroscience. *Trends in Cognitive Sciences*.
733 <https://doi.org/10.1016/j.tics.2018.01.005>
- 734 Fiedler, L., Wöstmann, M., Graversen, C., Brandmeyer, A., Lunner, T., & Obleser, J.
735 (2017). Single-channel in-ear-EEG detects the focus of auditory attention to

- 736 concurrent tone streams and mixed speech. *Journal of Neural Engineering*.
737 <https://doi.org/10.1088/1741-2552/aa66dd>
- 738 Fiedler, L., Wöstmann, M., Herbst, S. K., & Obleser, J. (2019). Late cortical tracking of
739 ignored speech facilitates neural selectivity in acoustically challenging conditions.
740 *NeuroImage*. <https://doi.org/10.1016/j.neuroimage.2018.10.057>
- 741 Hamilton, L. S., & Huth, A. G. (2018). The revolution will not be controlled: natural
742 stimuli in speech neuroscience. *Language, Cognition and Neuroscience*.
743 <https://doi.org/10.1080/23273798.2018.1499946>
- 744 Hasson, U., Nir, Y., Levy, I., Fuhrmann, G., & Malach, R. (2004). Intersubject
745 Synchronization of Cortical Activity during Natural Vision. *Science*.
746 <https://doi.org/10.1126/science.1089506>
- 747 Huk, A., Bonnen, K., & He, B. J. (2018). Beyond Trial-Based Paradigms: Continuous
748 Behavior, Ongoing Neural Activity, and Natural Stimuli. *The Journal of Neuroscience*.
749 <https://doi.org/10.1523/jneurosci.1920-17.2018>
- 750 Jessen, S., & Kotz, S. A. (2011). The temporal dynamics of processing emotions from
751 vocal, facial, and bodily expressions. *NeuroImage*, 58(2), 665–674.
- 752 Jones, E. J. H., Venema, K., Lowy, R., Earl, R. K., & Webb, S. J. (2015). Developmental
753 changes in infant brain activity during naturalistic social experiences. *Developmental*
754 *Psychobiology*. <https://doi.org/10.1002/dev.21336>
- 755 Kalashnikova, M., Peter, V., Di Liberto, G. M., Lalor, E. C., & Burnham, D. (2018). Infant-
756 directed speech facilitates seven-month-old infants' cortical tracking of speech.
757 *Scientific Reports*. <https://doi.org/10.1038/s41598-018-32150-6>
- 758 Lalor, E. C., Pearlmutter, B. A., Reilly, R. B., McDarby, G., & Foxe, J. J. (2006). The VESPA:
759 A method for the rapid estimation of a visual evoked potential. *NeuroImage*.
760 <https://doi.org/10.1016/j.neuroimage.2006.05.054>
- 761 Leong, V., Byrne, E., Clackson, K., Georgieva, S., Lam, S., & Wass, S. (2017). Speaker gaze
762 increases information coupling between infant and adult brains. *Proceedings of the*
763 *National Academy of Sciences*. <https://doi.org/10.1073/pnas.1702493114>
- 764 Maris, E., & Oostenveld, R. (2007). Nonparametric statistical testing of EEG- and MEG-
765 data. *Journal of Neuroscience Methods*, 164(1), 177–190.
- 766 Marshall, P. J., & Shipley, T. F. (2009). Event-related potentials to point-light displays of
767 human actions in 5-month-old infants. *Developmental Neuropsychology*.
768 <https://doi.org/10.1080/87565640902801866>

769 Matusz, P. J., Dikker, S., Huth, A. G., & Perrodin, C. (2018). Are we ready for real-world
770 neuroscience? *Journal of Cognitive Neuroscience*.
771 https://doi.org/10.1162/jocn_e_01276

772 Naselaris, T., Kay, K. N., Nishimoto, S., & Gallant, J. L. (2011). Encoding and decoding in
773 fMRI. *NeuroImage*. <https://doi.org/10.1016/j.neuroimage.2010.07.073>

774 Nishimoto, S., Vu, A. T., Naselaris, T., Benjamini, Y., Yu, B., & Gallant, J. L. (2011).
775 Reconstructing visual experiences from brain activity evoked by natural movies.
776 *Current Biology*. <https://doi.org/10.1016/j.cub.2011.08.031>

777 O'Sullivan, A. E., Crosse, M. J., Di Liberto, G. M., & Lalor, E. C. (2017). Visual Cortical
778 Entrainment to Motion and Categorical Speech Features during Silent Lipreading.
779 *Frontiers in Human Neuroscience*. <https://doi.org/10.3389/fnhum.2016.00679>

780 Oostenveld, R., Fries, P., Maris, E., & Schoffelen, J.-M. (2011). FieldTrip: Open source
781 software for advanced analysis of MEG, EEG, and invasive electrophysiological
782 data. *Computational Intelligence and Neuroscience*, 2011, 156869.

783 Pichon, S., de Gelder, B., & Grèzes, J. (2009). Two different faces of threat. Comparing the
784 neural systems for recognizing fear and anger in dynamic body expressions.
785 *NeuroImage*, 47, 1873–1883.

786 Picton, T. (2013). Hearing in Time. *Ear and Hearing*.
787 <https://doi.org/10.1097/aud.0b013e31827ada02>

788 Reid, V. M., Hoehl, S., & Striano, T. (2006). The perception of biological motion by infants:
789 An event-related potential study. *Neuroscience Letters*.
790 <https://doi.org/10.1016/j.neulet.2005.10.080>

791 Reynolds, G. D., & Guy, M. W. (2012). Brain-behavior relations in infancy: Integrative
792 approaches to examining infant looking behavior and event-related potentials.
793 *Developmental Neuropsychology*. <https://doi.org/10.1080/87565641.2011.629703>

794 Ringach, D., & Shapley, R. (2004). Reverse correlation in neurophysiology. *Cognitive*
795 *Science*. <https://doi.org/10.1016/j.cogsci.2003.11.003>

796 Ru, P. (2001). *Multiscale Multirate Spectro-Temporal Auditory Model*. University of
797 Maryland College Park.

798 Stets, M., Stahl, D., & Reid, V. M. (2012). A meta-analysis investigating factors underlying
799 attrition rates in infant ERP studies. *Developmental Neuropsychology*, 37(3), 226–252.
800 <https://doi.org/10.1080/87565641.2012.654867>

801 Vanrullen, R., & MacDonald, J. S. P. (2012). Perceptual echoes at 10 Hz in the human
802 brain. *Current Biology*. <https://doi.org/10.1016/j.cub.2012.03.050>

803 Wass, S. V., Noreika, V., Georgieva, S., Clackson, K., Brightman, L., Nutbrown, R., ...
804 Leong, V. (2018). Parental neural responsivity to infants' visual attention: How
805 mature brains influence immature brains during social interaction. *PLoS Biology*.
806 <https://doi.org/10.1371/journal.pbio.2006328>

807 Webb, S. J., Long, J. D., & Nelson, C. A. (2005). A longitudinal investigation of visual
808 event-related potentials in the first year of life. *Dev Sci*, 8(6), 605–616.
809 <https://doi.org/10.1111/j.1467-7687.2005.00452.x>

810 Wunderlich, J. L., & Cone-Wesson, B. K. (2006). Maturation of CAEP in infants and
811 children: A review. *Hearing Research*. <https://doi.org/10.1016/j.heares.2005.11.008>

812 Wunderlich, J. L., Cone-Wesson, B. K., & Shepherd, R. (2006). Maturation of the cortical
813 auditory evoked potential in infants and young children. *Hearing Research*.
814 <https://doi.org/10.1016/j.heares.2005.11.010>

815