

# 1 Prenatal development of neonatal vocalizations

2  
3 Darshana Z. Narayanan<sup>1,2</sup>, Daniel Y. Takahashi<sup>1,2\*</sup>, Lauren M. Kelly<sup>1,2</sup>, Sabina I. Hlavaty<sup>3</sup>, Junzhou  
4 Huang<sup>4</sup> and Asif A. Ghazanfar<sup>1,2,3\*</sup>

5  
6 <sup>1</sup>Princeton Neuroscience Institute

7 <sup>2</sup>Department of Psychology

8 <sup>3</sup>Department of Ecology & Evolutionary Biology

9 Princeton University

10 Princeton NJ 08544

11  
12 <sup>4</sup>Department of Computer Science and Engineering

13 The University of Texas at Arlington

14 500 UTA Blvd, ERB 650

15 Arlington TX 76019

16  
17 \*Correspondence: [takahashiyd@gmail.com](mailto:takahashiyd@gmail.com) or [asifg@princeton.edu](mailto:asifg@princeton.edu)

18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28

29 **Abstract**

30

31 Human and non-human primates produce rhythmical sounds as soon as they are born. These early  
32 vocalizations are important for soliciting the attention of caregivers. How they develop, remains a  
33 mystery. The orofacial movements necessary for producing these vocalizations have distinct  
34 spatiotemporal signatures. Therefore, their development could potentially be tracked over the course  
35 of prenatal life. We densely and longitudinally sampled fetal head and orofacial movements in  
36 marmoset monkeys using ultrasound imaging. We show that orofacial movements necessary for  
37 producing rhythmical vocalizations differentiate from a larger movement pattern that includes the  
38 entire head. We also show that signature features of marmoset infant contact calls emerge prenatally  
39 as a distinct pattern of orofacial movements. Our results establish that aspects of the sensorimotor  
40 development necessary for vocalizing occur prenatally, even before the production of sound.

41

42

## 43 Introduction

44  
45 Neonatal primates produce vocalizations (e.g., cries and coos) as a rhythmical series of sounds [1, 2].  
46 These vocalizations are often the substrate for the rhythmic vocalizations produced by adult humans  
47 [3] and other primates [4, 5](see [6] and [7] for reviews). Their production requires coordination  
48 among all the elements of the vocal apparatus; i.e., between laryngeal tension, respiration and  
49 orofacial movements. We know that postnatal vocal production is a complex motor act that changes  
50 over time and is influenced by caregiver feedback [8, 9] and the ambient environment [10]. But how  
51 do those very first neonatal vocalizations develop?

52  
53 During human pregnancies, movements consistent with crying have been observed during  
54 ultrasound assessments of fetuses in their last trimester [11] but their prenatal developmental  
55 trajectory and differentiation from other movements have not been measured. We reasoned that a  
56 nonhuman primate model of this behavior would be a way to gain insights. Marmoset monkeys are  
57 good candidate model for human infant vocal behavior. The early postnatal vocalizations of infant  
58 marmosets undergo a very similar developmental trajectory as infant human vocalizations: 1) They  
59 transition from noisy cries to tonal vocalizations that are adult-like [12]; 2) These early vocalizations  
60 are produced in bouts that have a rhythmic pattern that changes over time [13]; and 3) Vocal  
61 development is influenced by parental care and feedback [14-16]. These across-species similarities in  
62 postnatal vocal development are also occurring at the same life history stage (though marmosets  
63 develop 12 times faster than humans [17]), supporting the idea this species may also be a good  
64 model for prenatal human behaviors.

65  
66 We tracked the fetal orofacial movements of marmoset monkeys that potentially represent  
67 those movements necessary for their neonatal contact-calling. Neonatal marmoset contact  
68 vocalizations have a spatiotemporal signature distinct from any other vocalization in the marmoset's  
69 vocal repertoire. It therefore provided a template: We could compare prenatal mouth movements  
70 with those produced during neonatal contact calling. We quantified and characterized the prenatal  
71 trajectory of these movements to address the following questions: If and how do mouth  
72 movements differentiate from more global bodily movements, and if and how do mouth  
73 movements differentiate prenatally into those used for neonatal vocalizations?

## 74 Results

75  
76 We performed non-invasive ultrasound imaging on awake, pregnant marmoset monkeys (n=4  
77 pregnancies across two marmoset monkeys). Marmosets typically produce dizygotic twins at birth  
78 but not always [18]. In our sample, the 4 pregnancies consisted of one singleton, two sets of twins  
79 and one set of quadruplets. Since individual fetuses cannot be routinely identified via ultrasound,  
80 each pregnancy with more than one fetus were treated as a composite of a single subject. To track  
81 developmental trajectories, we densely sampled fetal orofacial movements from the day the face was  
82 clearly discernible to the day before birth (~embryonic day (E)95-146; 2-3 times per week with  
83 imaging sessions lasting between 15 and 45 minutes; 14-17 sessions per pregnancy). We then  
84 compared these movements to signature features of infant vocalizations in the first week after birth  
85 (postnatal day (PD) 1-7; n=7). We were thus able to observe the developmental trajectory of  
86 orofacial movements necessary for vocal production from the fetal to infant stage.

87  
88  
89 We had two basic hypotheses. The first was that fetal orofacial movements would initially be  
90 linked to movements of other body parts. Ultrasound studies of human fetuses support this

91 hypothesis, showing that isolated jaw movements appear late in development, weeks after the onset  
 92 of spontaneous general movements involving many body parts [19, 20]. Our second hypothesis was  
 93 that signature patterns of articulatory movements related to *postnatal* marmoset contact calls would  
 94 emerge *prenatally*. Investigations of the prenatal movements of rats, and pre-hatching movements of  
 95 birds, show that seemingly unorganized action patterns are followed by organized action patterns  
 96 that are reminiscent of postnatal or post-hatching movements. [21-27].  
 97

98 In each imaging session, we specifically targeted the face as our region of interest which  
 99 allowed us to track both orofacial movements and head movements. From these videos, we did a  
 100 frame-by-frame analysis (Fig. 1A). The developmental change was readily noticeable, and ostensibly  
 101 in keeping with our expectation that orofacial and head movements move together more often in  
 102 the younger compared to the older fetus (Fig. 1B-C; videos S1 and S2).

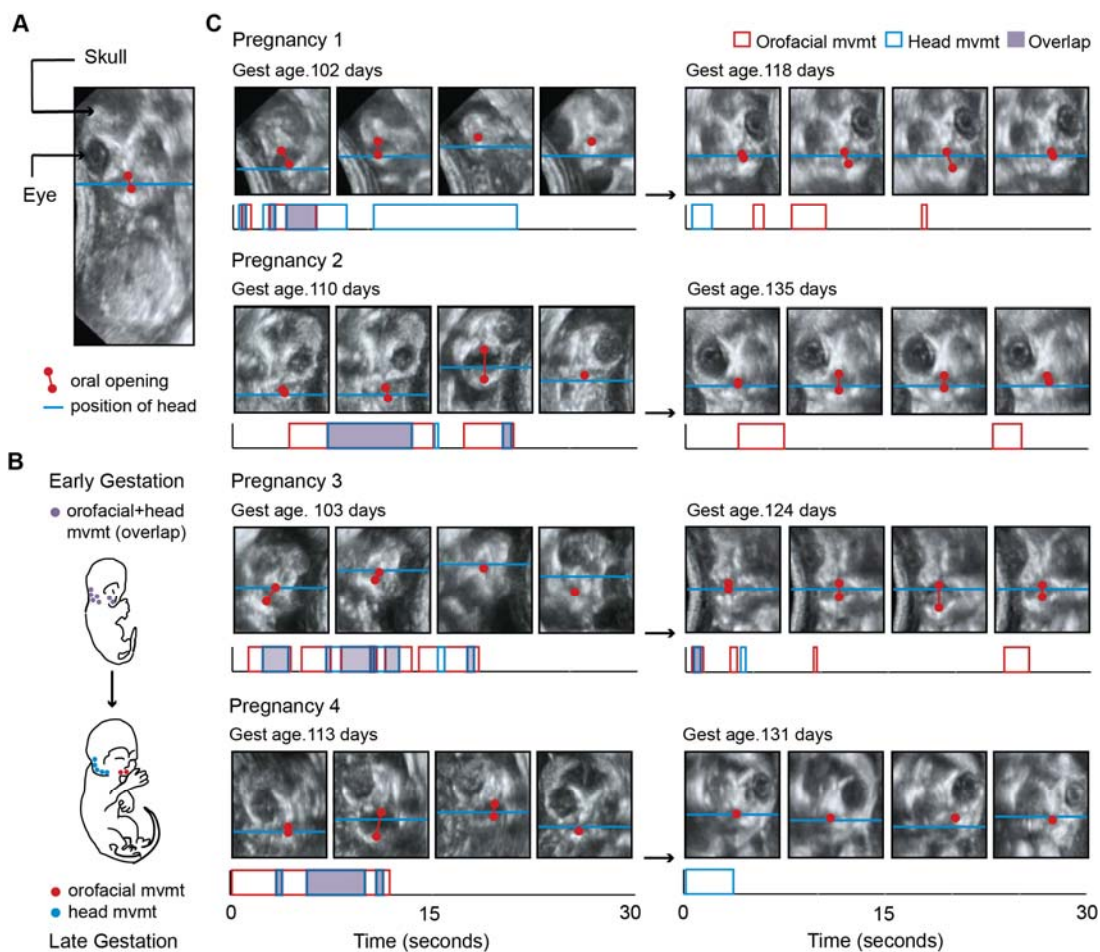


Fig 1. Fetal orofacial movements undergo striking changes through gestation, observable using ultrasound imaging. (A) Still frame from fetal ultrasound. Red is used to mark the upper and lower jaws and indicate oral opening. The blue line marks the position of the head, tracked using the jaw joint. (B) Illustration of the expected developmental change in orofacial and head movements. The top image depicts a young fetus, with orofacial and head regions often moving together. The bottom image illustrates a late-stage fetus, with orofacial and head regions moving in isolation. (C) Developmental change in orofacial and head movements for each pregnancy. Top: Still frames from representative ultrasound clips. Gestational day is indicated in the upper left of each panel. Below: Time stamps of the orofacial and head movements in the clip. Durations of orofacial (red) and head (blue) movements, and regions of overlap (purple) are indicated. The amplitude of movement has no bearing on our analysis. This developmental change can be observed by comparing videos S1 and S2.

103 One of the hallmarks of development is increasing ‘order’ in behavior through self-  
104 organization. Since the relationship between fetal head and orofacial movements in marmosets  
105 changed over time, we tested if, for these two movement behaviors, there is concomitant increase of  
106 structure in their action patterns, by mapping the sequence of five different states: independent  
107 orofacial movements (State 1); independent head movements (State 2); orofacial movements  
108 followed by overlapping head movements (State 3); head movements followed by overlapping  
109 orofacial movements (State 4); and synchronous orofacial and head movements (State 5). The fetal  
110 behavioral states are seen to be more numerous and variable in the early weeks of gestation, when  
111 compared to late gestation. Figure 2A shows the proportion of states through gestation for the  
112 population data; Figure 2B shows the state diagrams for exemplar sessions selected from different  
113 gestational periods of one pregnancy.

114  
115 We quantified these state changes in two ways. First, Shannon entropy was used to measure  
116 behavioral variability within each session (Fig. 2C). For determining the developmental trajectory,  
117 the best polynomial-fit order using AIC was found to be three. Behavioral variability decreased  
118 through gestation and then remained steady at  $\sim 0.8$  bits (the maximum entropy for a behavior with  
119 5 possible states is 2.32 bits). The decrease in entropy ( $p < 0.001$ , test of nullity of the relation  
120 between gestational day and entropy) indicates that as the fetus gets older, there is increasing  
121 movement structure.

122  
123 Second, we performed a Kullback-Leibler divergence test to quantify the behavioral change  
124 through fetal life (Fig. 2D). The state distribution on the earliest imaging day was compared with the  
125 state distributions of all later imaging days. The best polynomial-fit order for the divergence  
126 estimates was one. The resulting linear fit with a positive slope ( $p < 0.001$ ) indicates that with  
127 increasing gestational age, fetal behavior — with respect to orofacial and head movements —  
128 becomes increasingly different from the first imaging day of orofacial movement. Thus, what we see  
129 is that in the young fetus, orofacial and head body parts move in a number of different ways —  
130 independently, and also often together in various combinations (orofacial-head, head-orofacial or  
131 synchronous onset), but when close to birth, orofacial and head body parts are, with rare exception,  
132 moved separately. This suggests that these two motor regions are decoupled in the late-stage fetus,  
133 which allows the newly born infant to use head and orofacial movements adaptively; in marmoset  
134 infants, this would putatively be for the different purposes of orienting, feeding and vocalizing.

135  
136 To further quantify the developmental change, we examined the profiles of orofacial  
137 movements ( $n=1977$ ) and head movements ( $n=1216$ ) independently. We calculated the rates of  
138 orofacial and head movements (number of movements per hour) in each imaging session. The best  
139 polynomial-fit order was two for the orofacial movements, and one for head movements. The  
140 occurrence of orofacial movements showed an inverse U-shaped profile; steadily increasing from  
141  $\sim E93-105$ , hitting a peak from  $\sim E106-131$ , and declining from  $\sim E119-131$  (Fig. 2E, in red). The  
142 head movements, however, showed a linear trend. For both orofacial and head movements, the  
143 same patterns were observed in the individual pregnancies (Fig 2F; clockwise from top left:  
144 singleton, quadruplets, twins, twins).

145  
146 To study the putative “linked” relationship between these two motor patterns, we calculated  
147 the percentage of temporal overlap between them (total instances of overlap divided by the total  
148 number of movements). A first order polynomial curve was found to be the best fit for determining  
149 the overlap profile. We found that the overlap of orofacial and head movements decreases over the  
150 course of fetal development (Fig. 2G in purple,  $p < 0.001$ ). One possibility is that, since movements

151 decrease in general towards the end of gestation, the decrease in overlapping movements simply  
 152 becomes less probable. To account for this, we did a permutation test to be certain the decline in  
 153 overlapping movements was independent of general movement decline. Orofacial and head  
 154 movements within each session were independently shuffled, keeping the duration and latency  
 155 distributions intact. Percent overlap was calculated for the permuted data, and a first order  
 156 polynomial curve was fitted. The permutation test generated a flat profile ( $p=0.057$  for the  
 157 significance of the mean regression line) compared to the overlap decline seen in the actual data,  
 158 confirming that the decline in overlap is not simply due to the decline in overall movement quantity  
 159 (Fig. 2G). Another possibility is that movement duration decreases throughout gestation which  
 160 would therefore decrease the probability of overlapping movements. This was not the case: The  
 161 durations of mouth movements increased with time ( $p=.024$ ) and the durations of head movements  
 162 remained steady throughout gestation ( $p=.093$ ). Thus, consistent with our prediction that mouth  
 163 movements would be linked to other body parts, orofacial and head movements are coupled early in  
 164 gestation but become increasingly independent. This held true for each of the four pregnancies (Fig.  
 165 2H; clockwise from top left: singleton, quadruplets, twins, twins).

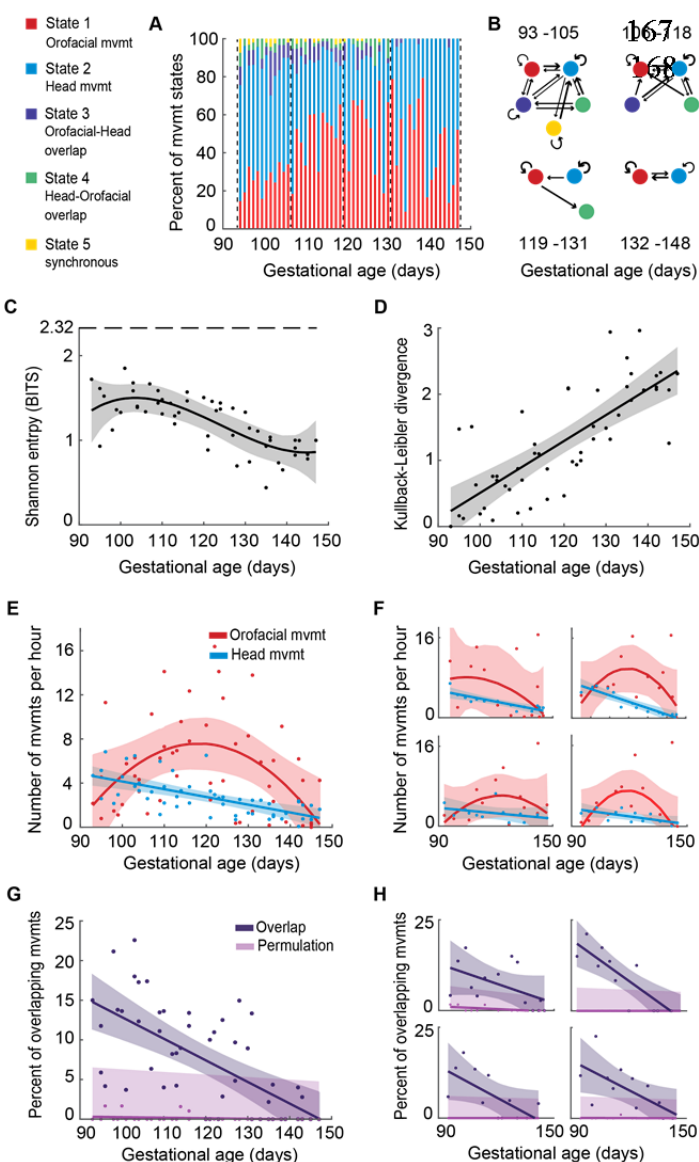


Fig 2. Fetal orofacial movements differentiate from a larger movement pattern that includes the head, leading to an increase in ‘order’ in the motor behavior of the fetus. (A) Proportion of behavioral states per session. A total of 5 states are represented in red (isolated orofacial movements), blue (isolated head movements), purple (overlap: orofacial followed by head), green (overlap: head followed by orofacial), and yellow (synchronous movements). (B) Transition diagrams visualizing the behavioral states in four sessions chosen from different stages of gestation of a single pregnancy. The widths of the arrows indicate the transition frequencies between states. (C) Developmental change in the behavioral variability per session. Points represent entropy measures for single sessions, and the curve represents optimal polynomial fit based on AIC. 2.32 bits is the maximum entropy for a behavior with 5 possible states. (D) Kullback-Leibler (KL) divergence of behavioral state distributions. Points represent the relative entropy measures for every session compared to the first testing session, and the curve represents optimal polynomial fit. (E) Developmental change in the rates of orofacial and head movements. Points represent the rate of orofacial (red) and head (blue) movements in each session, the curves represent the optimal polynomial fits. (F) Same as (E), for each pregnancy. (G) Developmental change in overlap of orofacial and head movements. Points represent the percentage of overlap in each session, curves represent optimal polynomial fits, shaded regions denote the 95% confidence interval for the fits. (H) Same as (G), for each pregnancy.

169 Regarding vocalizations, we next focused only on the patterns of fetal orofacial movements  
170 to test the hypothesis that they gradually exhibited the temporal organization of the infant marmoset  
171 monkey's rhythmic contact calling [13]. To do this, we quantified the temporal pattern of contact  
172 calling by neonatal marmosets (within ~24 hours of birth); these were the same individuals that were  
173 imaged as fetuses (n=7; one singleton, two sets of twins, two surviving individuals from the  
174 quadruplets). Contact calling was induced by briefly separating infants from caregivers; we acquired  
175 ~5 minutes of audio recordings and ~5 minutes of both video and audio. In both media, infant  
176 contact calls are easily identifiable by their long duration and many syllables[12, 13] (Fig 3A). We  
177 compared the recordings of infants producing contact calls to imaging data from fetuses producing  
178 orofacial movements (Fig 3B shows stills from the videos of an infant at P1 and a fetus at E118).  
179 The upper and lower jaws were tracked frame-by-frame to generate temporal profiles of the  
180 movements. We found that a subset of the fetal orofacial movements was a match to the infant  
181 contact call in terms of overall temporal profile, duration and syllable number. (Fig 3C-F; video S3).  
182

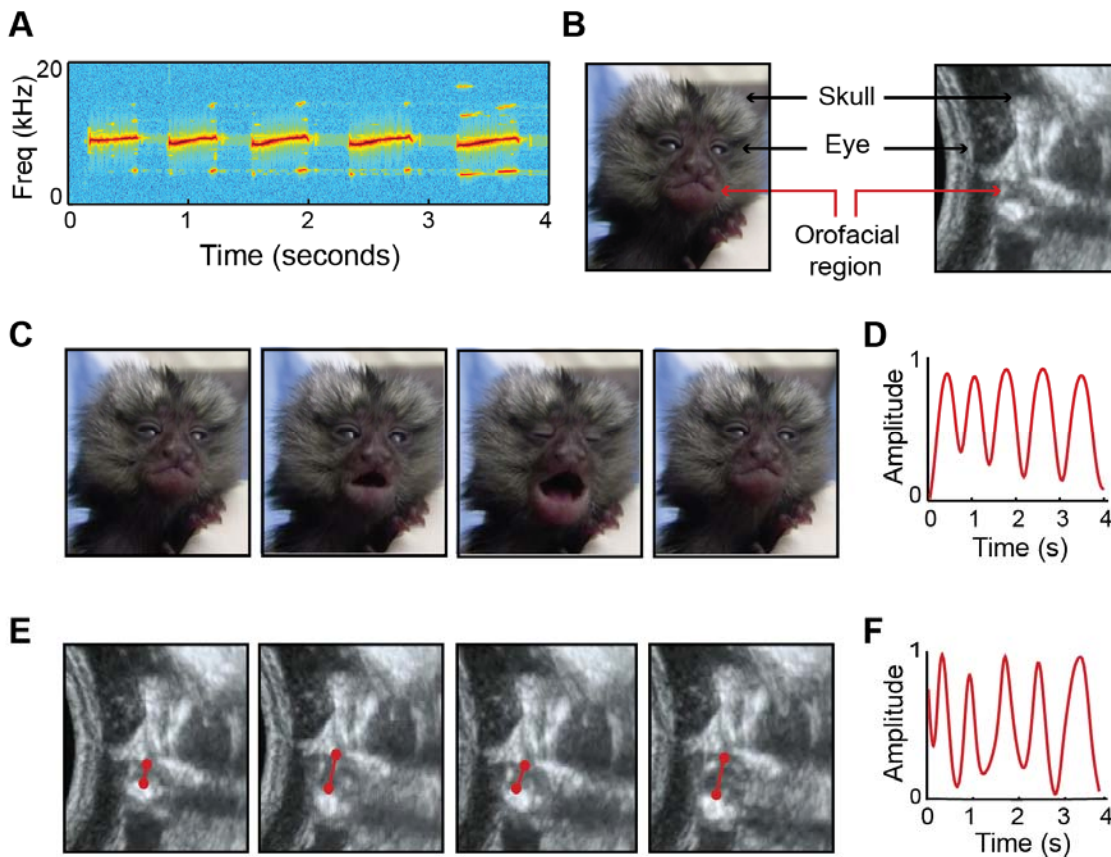


Fig 3. Signature features of marmoset infant calls observed in fetal movements. (A) Spectrogram of a five-syllable infant contact call from postnatal week 1. (B) Stills of marmoset infant and fetus, indicating orofacial region (C) Video stills of week 1 infant producing a five-syllable contact call. (D) Temporal profile of the infant contact call in (C), generated by tracking orofacial movements. (E) Ultrasound stills of the orofacial movements of a late-stage marmoset fetus. (F) Temporal profile of the fetal orofacial movements in (E). The match between infant and fetal orofacial movements can be observed in video S3.

183  
184

185 The cyclic nature of the infant contact call is fully captured in its temporal profile. We  
186 therefore matched the temporal profiles of fetal orofacial movements to those of infant contact  
187 calls. To do this, we took the video recordings (n=25) of neonatal marmoset infants producing  
188 contact calls and did a frame-by-frame tracking of the upper and lower jaw positions to generate  
189 contact call profiles. Multiple contact call profiles of the same syllable number were averaged and  
190 smoothed to create a set of contact call ‘templates’ (n=7; 1-7 syllables). To get the temporal profiles  
191 of fetal orofacial movements, we used a custom-made computer vision program that tracked each  
192 movement. Only temporal profiles that matched the manually calculated syllable number features of  
193 the movement were included in the analysis (n=414). We smoothed these profiles using a Savitzky-  
194 Golay Smoothing Filter (polynomial degree=3; sliding window size=9). We then used dynamic time  
195 warping (DTW) analysis [28] to measure the similarity between the temporal profiles of the fetal  
196 orofacial movements and the infant contact call templates. Every fetal orofacial movement profile  
197 was compared with every contact call template, and the one with the smallest DTW cost (i.e., the  
198 closest similarity) was chosen. The median DTW cost was calculated for each session (n=57). Our  
199 prediction was that the DTW cost would decrease through gestation, indicating that the temporal  
200 profiles of the fetal orofacial movements are getting closer to those of infant contact calls (Fig 4A).  
201 We found this to be the case (Fig 4B). A multiple linear regression analysis controlling for the  
202 different pregnancies shows that gestational age predicts the prevalence of ‘contact call-like’  
203 temporal profiles ( $\beta \pm SE = -0.005 \pm 0.003$ ,  $t = -1.97$ ;  $F(1,56)=4.60$ ;  $p=.037$ ). In all four pregnancies,  
204 fetal orofacial movements increasingly match contact call temporal profiles as gestation progresses  
205 (Fig 4C; clockwise from top left: singleton, quadruplets, twins, twins).  
206

207 We further quantified the developmental change leading up to this match in a number of  
208 ways. We first looked at the durations of fetal orofacial movements over the course of gestation  
209 (n=1845; Fig 4D). The median, 75th percentile, and 25th percentile durations were calculated for  
210 each session. To determine the developmental trajectory, the best polynomial-fit order using AIC  
211 was found to be three for the median values. The same polynomial order was used for generating  
212 curves for the 75th and 25th percentiles values. Kernel density estimation was used to visualize the  
213 duration density change over the course of gestation. The durations of orofacial movements show  
214 an increasing trend. Comparing the fetal orofacial movement durations to those of P1 infant contact  
215 calls, the fetal movements are seen to increasingly approach the P1 durations (Fig 4E; n=120;  
216 median duration=3.69s; SE=0.14).  
217

218 Since contact calls typically contain multiple syllables, we calculated the ‘syllable number’ of  
219 the fetal orofacial movements (n=1845; Fig 4F). This is the number of individual orofacial  
220 movements separated by <500ms (i.e., the same criteria used for counting actual contact call  
221 syllables in infants [13]). The median, 75th percentile, and 25th percentile syllable numbers were  
222 calculated for each session. The best polynomial-fit order using AIC was found to be two for the  
223 median values. The same polynomial order was used to fit curves for the 75th and 25th percentiles  
224 values. The polynomial curves and kernel density estimation show that the syllable numbers of fetal  
225 orofacial movements increase through gestation. Fetal syllable numbers get closer to the syllable  
226 numbers of contact calls produced on P1 (Fig 4G; n=120; median=5; SE=0.18).  
227

228 To directly compare the features of fetal orofacial movements with signature features of the  
229 infant contact call, we established criteria for identifying the contact call based on duration and  
230 syllable number. Contact calls tend to have long durations, and multiple syllables when compared to  
231 other infant calls. An infant call within the duration range of 3.69–6.5s and syllable number range 5-  
232 9 has a 97% likelihood of being a contact call. To track the development of ‘contact call-like’ fetal



233 orofacial movements, we calculated, for each session, the percentage of movements that matched  
 234 the duration and syllable number features of P1 contact calls (Fig 4H). A multiple linear regression  
 235 analysis controlling for the different pregnancies shows that fetal orofacial movements increasingly  
 236 match the contact call profile as the fetus gets older ( $\beta \pm SE = 0.15 \pm 0.04$ ,  $t = 3.44$ ;  $F(1, 63) = 12.09$ ,  
 237  $p = .001$ ). All four pregnancies were seen to have more ‘contact call-like’ movements over the course  
 238 of gestation (Fig 4I; clockwise from top left: singleton, quadruplets, twins, twins).  
 239

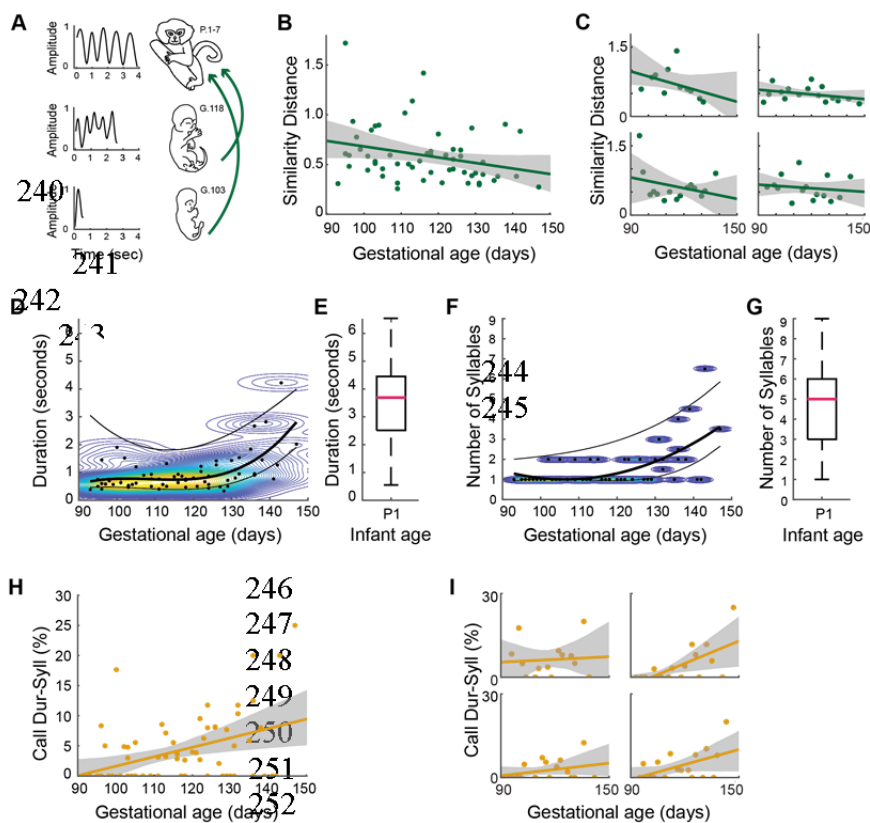


Fig 4. Signature features of marmoset infant calls on P1-7, emerge prenatally as distinct patterns of orofacial movements. (A) Illustration of the gestational change in the temporal profiles of fetal orofacial movements. On the left are the temporal profiles generated by tracking orofacial movements. On the right are the ages of the animals. (B) Matching temporal profiles of fetal orofacial movements to that of week 1 infant contact calls. Points represent the DTW similarity distance (lower the similarity distance, higher the matching), and the line represents the regression fit. (C) Same as (B), for each pregnancy. (D) Gestational change in the durations of fetal orofacial movements. Points represent the median movement duration per session. The curves are fit to the median, 75 percentile, and 25 percentile values. The background contour map indicates durations with the highest density of movements (red: high density; blue: low density). (E) Infant contact call durations on P1. The pink bar marks the median contact call duration. (F) Gestational change in the syllable number of fetal orofacial movements. Points represent the median syllable number per session. The curves are fit to the median, 75 percentile, and 25 percentile values. The curves were generated by first calculating the optimal polynomial degree for the median values. The background contour map indicates syllable numbers with the highest density of movements (yellow: high density; blue: low density). (G) Infant contact call syllable numbers on P1. The pink bar marks the median syllable number. (H) Fetal orofacial movements matching infant contact call duration + syllable number signatures. Points represent the match percentage per session, and the line represents the regression fit ( $p = .001$ ). (I) Same as (H), for each pregnancy.

253 To control for the possibility that the fetal orofacial  
 254 movement profile is a generic one not specifically linked to contact  
 255 calls, we compared it to the two other orofacial movements common  
 256 in the first weeks of postnatal life: twitter vocalizations and licking.  
 257 When separated from caregivers, marmoset infants also produce  
 258 bouts of another type of vocalization known as twitters. Relative to  
 259 contact calls, twitters are short duration calls with fewer syllables.  
 260 First, we checked if this contrasting type of vocalization produced by the P1 infants ( $n=109$  twitters)  
 261 follows a similar developmental trajectory. An infant call within the duration range of 0.96–1.5s and  
 262 syllable number range 2-3 has a 86% likelihood of being a twitter call (and a 2% likelihood of being a  
 263 contact call). We then calculated, for each session, the percentage of fetal movements that matched  
 264 the duration and syllable number features of P1 twitter calls (Fig 5A). A multiple linear regression  
 265 analysis controlling for the different pregnancies reveals that — in contrast to what is seen with the  
 266  
 267  
 268  
 269

270 contact call — fetal orofacial movements do not increasingly match the twitter profile as the fetus  
 271 gets older ( $\beta \pm SE = 0.06 \pm 0.06$ ,  $t = 0.99$ ;  $F(1, 63) = 0.87$ ,  $p = 0.35$ ).

272  
 273 We observed that when separated from their caregivers the infants often produced licking  
 274 movements. [Note: Weaning begins at  $\sim 30$  days in marmosets, therefore chewing movements are  
 275 not seen in neonates.] We measured the duration and syllable number features of these movements  
 276 in the same infants, from ages P1-7 ( $n = 37$ ) and found that licks have greater variability than the  
 277 contact call both in duration (Fig 5B; median = 1.6s;  $SE = 0.83$ ) and “syllable” number (number of jaw  
 278 movements) (Fig 5C; median = 2;  $SE = 0.85$ ). We found a 0% likelihood that a movement within the  
 279 duration range of 3.69–6.5s and syllable number range of 5-9 (our contact call criteria) is a licking-  
 280 related movement (Fig 5D). Infant licking movements are distinct from their contact calls. We then  
 281 tracked the development of ‘lick-like’ orofacial movements in the fetus. The licks of P1-7 infants  
 282 tend to be short with few syllables, or very long with a high number of syllables. To capture the  
 283 maximum possible infant licks, we established separate criteria for each of these types of licks:  
 284 duration range of 0.33–3.77s and syllable number range of 1-6 for the short licks, and duration range  
 285 of 6.67–14.29s and syllable number range of 10-16 for the long licks. Together they capture 83.78%  
 286 of the infant licks (and 0% of the infant contact calls). We then calculated, for each session, the  
 287 percentage of fetal movements that matched the features of licks (Fig 5E). We identified more ‘lick-  
 288 like’ orofacial movements than ‘contact call-like’ movements in the fetus, but a multiple linear  
 289 regression analysis controlling for the different pregnancies reveals that fetal orofacial movements  
 290 do not increasingly match the lick profile as the fetus gets older ( $\beta \pm SE = 0.10 \pm 0.11$ ,  $t = 0.97$ ;  $F(1,$   
 291  $63) = 0.78$ ,  $p = 0.38$ ). These comparisons establish that the changes in movements we have tracked are  
 292 specifically related to the rhythmic contact calls made by neonatal marmosets. As we hypothesized,  
 293 the marmoset contact call production has a period of prenatal development.  
 294

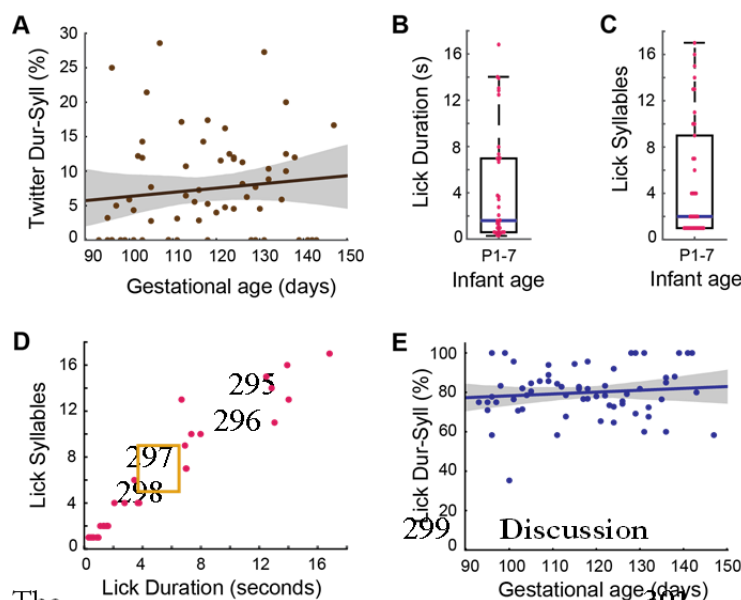


Fig 5. Prenatal developmental change specific to contact calls (A) Fetal orofacial movements matching infant twitter call duration + syllable number signatures. Points represent the match percentage per session, and the line represents the regression fit ( $p = 0.35$ ). (B) Infant lick durations on P1-7. Points represent the durations of individual movements. The blue bar marks the median lick duration. (C) Infant lick call syllable numbers on P1-7. Points represent the syllable numbers of individual movements. The blue bar marks the median syllable number. (D) Infant lick call duration + syllable number signatures distinct from contact calls. Points represent the duration and syllable number features of individual licks. The yellow box represents the contact call duration and syllable number profile (E) Fetal orofacial movements matching infant lick call duration + syllable number signatures. Points represent the match percentage per session, and the line represents the regression fit ( $p = 0.38$ ).

300  
 301 The neonate’s well-being hinges on using  
 302 vocalizations to solicit attention. In human and non-human primates, these neonatal vocalizations  
 303 transform into mature vocal signals through continuous and reciprocal interactions among the  
 304 infant’s advancing vocal apparatus, neural circuitry and interactions with caregivers [29-31]. Our  
 305 study focused on the prenatal origins of those very first vocalizations; particularly the development  
 306 of their spatiotemporal signatures, which we tracked using vocalization-related mouth movements.

307 Using fetal marmoset monkeys, we show that early mouth movements are tightly coupled to head  
308 movements. As gestation progresses, the movements decouple, ultimately leading to the formation  
309 of two independent movements. In parallel, a subset of fetal orofacial movements gradually assumes  
310 the pattern necessary to produce the immature-sounding contact calls neonates use to solicit care  
311 [32] and the parental interactions used to ratchet their vocal development [12, 14-16, 33]. Aspects of  
312 the sensory-motor development necessary for vocalizing are occurring prenatally, even before the  
313 production of sound.

314  
315 That body parts (head and mouth movements) were initially linked in marmoset fetuses is  
316 consistent with the developmental pattern of fetal limb movements across many species [25, 34-37].  
317 For example, early in gestation, embryonic chicks move their limbs and wings together. It is only in  
318 later pre-hatching life that these two body parts uncouple in their movements and can perform  
319 separate functions [24]. In sheep, the first limb movements are paired with head and neck motions,  
320 but these different body parts acquire increasing autonomy over the course of development [34].  
321 These fetal limb movements play an important role in the formation and organization of the spinal  
322 cord and central nervous system [38]. However, the development of these head and mouth action  
323 patterns in marmosets involve central pattern generators located in the brainstem not the spinal cord  
324 [39, 40]. This suggests that the principles of fetal development--whereby linked movements  
325 involving multiple body parts are gradually unlinked--are general.

326  
327 One question is whether the underlying mechanisms are the same for the patterns of change  
328 seen for head-mouth movement differentiation versus the differentiation between limb movements  
329 and other parts of the body. Studies of chicks suggest that the constraints of the egg help  
330 differentiate motor units and construct species-typical patterns of embryonic motility [41].  
331 Increasing uterine constraint has also been used to explain the common inversed-shaped pattern of  
332 limb movement frequency observed in many species [34, 36, 37, 42, 43]. In our study of marmosets,  
333 and in humans [43], the frequency of mouth movements over the course of gestation also showed  
334 an inversed U-shaped trajectory. Do they decrease in number because, as the fetus grows larger,  
335 there is less room to make orofacial movements? This seems unlikely in our study. In the four  
336 pregnancies we monitored, the number of fetuses (and thus the space constraints) varied, but there  
337 was no indication that there was any related difference in the developmental trajectory of mouth  
338 movements. Consistent with this idea, a study of fetal rats compared their activity levels under three  
339 different, decreasing space constraints: the uterine environment, outside the uterus but with extra-  
340 embryonic membranes intact, and with all membranes removed [23]. In all cases, activity profiles still  
341 exhibited the familiar inversed U-shaped profile [23]. These data suggest that the shape of the  
342 developmental trajectory is influenced by neural changes, possibly from the developing forebrain  
343 [34, 44]. Indeed, a computational model of motor development supports the idea that spontaneous  
344 motor activity can self-organize into reflex actions, and that these actions are subsequently  
345 modulated to produced coordinated, adaptive behaviors [45].

346  
347 *In utero* interactions between the species-typical body and brain, and between the body and  
348 its environment, are hypothesized to drive the development of muscle and neural organization, and,  
349 ultimately, lead to the development of species-typical behavior [41]. This idea is supported by  
350 computer simulations of fetal development that model fetal motor behavior and spinal circuit  
351 changes of different mammalian species within a uterine environment [42-45]. At beginning of  
352 simulations, all fetuses had different musculoskeletal bodies but the same undifferentiated neural  
353 circuits. Following spontaneous activity, the simulated fetuses developed species typical motor  
354 patterns, e.g., the human fetus developed alternating leg movements and the zebrafish embryo

355 developed coordinated side-to-side movements. A by-product of these simulations was that  
356 individual muscle contractions were initially part of larger undifferentiated motor units (e.g., a single  
357 unit controlling the head and trunk), but as development progressed, these large muscle units  
358 subdivided into smaller, more precisely controlled units (e.g., separate control of head and trunk).  
359 Our finding that head and mouth motor units are linked early in gestation and uncouple with age  
360 provide additional support for this hypothesis.

361  
362 A subset of fetal marmoset mouth movements were precursors to the movements needed to  
363 produce contact calls as neonates; over gestation, they became increasingly similar in their temporal  
364 structure to neonatal articulatory movements. Likewise, the movement patterns of fetal rodent limbs  
365 are seen to get quantitatively and qualitatively closer to those of the neonate [22, 46, 47]. These same  
366 phenomena are observed in chicken embryos [48, 49]. These studies in rodents and birds suggest  
367 that experiences in-utero play a key role in establishing early postnatal movements. For example,  
368 fetal rats modify their limb movements in response to physical constraints, exhibiting motor learning  
369 *in utero* [47]. The simulations described above also support the notion that *in utero* learning via  
370 sensory feedback generated by fetal movements lead to species-typical stepping or swimming actions  
371 [50-55]. Thus, it does not seem too far-fetched to suggest that self-generated experience leads to the  
372 convergence of the temporal structure of marmoset fetal mouth movements towards the pattern of  
373 neonatal vocal articulatory movements (at least for contact calls). Self-generated learning of this type  
374 would be of the same kind that is observed postnatally in mammals, whereby myoclonic twitches  
375 during sleep help construct sensorimotor maps in the brain [56-58](for review, see [59]).

376  
377 Here are some caveats to our study. Our study focused only on contact calls because they  
378 are among the first vocalizations produced by infants (albeit immature sounding) and are important  
379 for soliciting both vocal responses [33] and retrieval from caregivers [32]. The most obvious parallel  
380 to marmoset monkey contact calling (and likely all infant contact calling in mammals and birds) is  
381 crying by human infants. The comparison, however, is not so simple. The contact vocalizations that  
382 infant marmosets use to solicit caregiver attention are also used by adults to maintain social contact  
383 with conspecifics; this type of vocal behavior seems to be true for all nonhuman primates [60]. In  
384 humans, communication with caregivers through crying soon gives way to other types of  
385 vocalizations and eventually speech [61]. Next, in pregnancies with multiple fetuses, we treated them  
386 as a single composite subject. This was because there was no way to identify the same individuals  
387 across days. Two findings mitigate against this issue: all four pregnancies exhibited the same  
388 behavioral profiles and developmental trajectories, and the early postnatal vocal output of twins are  
389 nearly identical in their patterns of output (more so than in comparison to the their non-twin  
390 siblings), seemingly driven by arousal levels regulated by their caregivers [13]. Another caveat is that  
391 while we implicitly interpreted the fetal orofacial movements we observed as being “spontaneous” as  
392 other similar studies have, it is possible that these movements could have been initiated by activation  
393 of the developing sensory system [25]. However, we did not observe any correlation with, for  
394 example, application of the ultrasound probe and fetal movements, and a pilot study by us to elicit  
395 fetal movements via sound playback near the mother’s belly was unsuccessful. Finally, our sample  
396 size consisted of four pregnancies across two unrelated mothers. The imaging and analysis were very  
397 time intensive, and we opted to longitudinally sample individual pregnancies very densely as  
398 opposed to increasing the number of subjects. This allowed us to better map the shape of  
399 developmental trajectories [63].

400

401 Marmoset monkeys do offer a special opportunity for comparative investigations of vocal  
402 development with humans. Like humans, marmosets exhibit a high degree of prosociality [64] (and  
403 related phenotypes linked to “self-domestication” [65]), show context-dependent vocal control and  
404 plasticity in the timing, intensity and spectral features of their vocalizations [66-69], and infant  
405 marmosets undergo a period of vocal learning that is influenced by parents [12, 14-16]. All of these  
406 similarities may be related to the fact that both humans and marmosets are cooperative breeders and  
407 are born altricial relative to other primates [70]. Thus, we predict that the prenatal patterns of  
408 vocalization-related orofacial movements in marmosets we observed here is of the same kind in  
409 humans. This would open up possibilities to use marmoset monkeys to investigate how very early  
410 prenatal orofacial motor behaviors can be diagnostic of later vocal behaviors and prelinguistic vocal  
411 learning.

412

413

414

415

416  
417  
418  
419  
420  
421  
422  
423  
424  
425  
426  
427  
428  
429  
430  
431  
432  
433  
434  
435  
436  
437  
438  
439  
440  
441  
442  
443  
444  
445  
446  
447  
448  
449  
450  
451  
452  
453  
454  
455  
456  
457  
458  
459  
460  
461  
462

## Materials and Methods

### *Subjects*

The subjects used in this study were 9 fetuses from 4 pregnancies; 3 pregnancies of the same adult female, and 1 pregnancy of a different adult female. The two pregnant females came from different social groups and lineages. Of the four pregnancies, one was of a singleton, two of twins and one of quadruplets. Of the nine fetuses, seven survived (two of the quadruplets were stillborn). There were no obvious differences in the sizes of the singleton, twins, or quadruplets at birth. Our veterinary doctor performed necropsies on the stillborn animals and concluded that both were well developed and looked to be at term. The seven surviving fetuses were then tested as infants between postnatal days 1 and 7. The colony room was maintained at a temperature of  $\sim 27$  °C with 50–60% relative humidity and a 12:12 light/dark cycle. The animals had *ad libitum* access to water and were fed daily with a commercial marmoset diet, supplemented with fresh fruits, vegetables, and insects. All experiments were performed with the approval of the Princeton University Institutional Animal Care and Use Committee (protocol # 1908-18).

### *Experimental setup – ultrasonography*

Ultrasonography tests were performed using a GE Voluson i Ultrasound machine. The testing method used was adapted from the procedure developed by S.D. Tardif and colleagues [71]. Each examination was conducted by two experimenters — one gently restrained the animal while the other carried out the ultrasound procedure. The animals were not anesthetized for the examination. Instead, they were trained to accept restraint by hand and the gentle placement of the probe on their stomach. They were rewarded with treats at the end of the testing period. We routinely scanned all our adult females for pregnancy. Early pregnancy was detected by assessing the morphology of the uterus [71]. Once pregnancy was confirmed, ultrasounds were conducted every two weeks to monitor the development of the fetus. When the fetal skull became clearly visible, gestational age could be estimated by measuring the biparietal diameter of the skull [71]. At this point, we started monitoring the fetus 2-3 times a week. As soon as the fetal face was clearly visible ( $\sim 95$  days gestational age), the ultrasound imaging was fixed on the face as much as possible, and recordings were taken for 15-45 min every other day (by this time the mothers had become accustomed to the procedure, and we saw no evidence in their behavior that their stress levels systematically increased through the gestation period). In cases where there was more than one fetus, two fetuses were selected at random, and each fetus was observed for roughly half the time. We differentiated the fetus by its position in the womb during the session. The examination was terminated if the animal showed significant resistance, hence the variability in session length. All examinations were conducted between 1400 and 1800 hours, and the procedure was repeated using exactly the same procedure until birth ( $\sim 146$  days gestational age). The ultrasound videos were captured at a frame rate of 30Hz and written on to DVD for later analysis. A total of 64 sessions were recorded – 14 sessions for the singleton, 17 sessions for the first set of twins and the quadruplets, 16 sessions for the second set of twins.

463  
464  
465  
466  
467  
468  
469  
470  
471  
472  
473  
474  
475  
476  
477  
478  
479  
480  
481  
482  
483  
484  
485  
486  
487  
488  
489  
490  
491  
492  
493  
494  
495  
496  
497  
498  
499  
500  
501  
502  
503  
504  
505  
506  
507  
508  
509

### *Identifying orofacial and head movements*

To reduce experimenter error we developed explicit criteria for scoring the videos, and roughly half of the total 64 sessions were coded by one experimenter, and the other half by another (every other session done by the same person). Randomly chosen recording session videos were coded by both experimenters to check for coding consistency. The marmoset mothers were not anesthetized for the ultrasound procedure, and both the mother and the fetus could move, causing us to lose sight of the fetal face from time to time. Therefore, we first identified segments of the video during which the fetal face was clearly visible. Within these segments, we scored for orofacial and head movements.

Video analysis was done using Adobe Premiere Pro software, which enabled frame-by-frame analysis of the ultrasound movement segments. The ultrasound still images are visually similar to x-ray images. The bones of the face are seen in white, while the mouth and eye cavities appear dark in contrast. A movement counted as an orofacial movement when there was a clear separation of the upper and lower jaws. The first video frame where the jaws separated, and the dark region between the upper and lower jaws first began to increase, counted as the beginning of orofacial movement. The video frame where the jaws fully came back together was the end point of the movement. If we lost focus of the face before the jaws came back together, the last frame of observation counted as the end of the movement. We did not consider the surface area of mouth opening or the amplitude of the orofacial movements in our analysis as those measures are sensitive to the position and orientation of the moving fetus. We therefore only considered the relative position of the upper and lower jaws, which is easily and consistently discernible, even in fetuses as young as E90.

Individual orofacial movements were considered to be part of the same movement unit if they were separated by 500ms or less (15 frames). The 500ms criterion was justified by the bimodal structure of the inter-syllable interval distribution of the vocal output of marmoset neonates. In their calls, the 500ms threshold separates the first mode of the distribution (representing the interval between syllables within a single call) from the second mode that represents the interval between the offset of the last and onset of the first syllables between two calls [13]. We made a note of partially captured movement units. These units were included when counting the number of movement units, as in Fig 2. We excluded the partially captured movement units when calculating duration and syllable number, as in Fig 4, since the fetus moved out of focus while the mouth was still open.

Similar to orofacial movements, the first video frame where the head moved away from baseline position was the beginning of movement and the first video frame where the head came back to baseline was the end point of movement. The baseline position of the head was determined using the region where the upper and lower jaw meet, so as not to conflate head movements with the movement of the lower jaw. If we lost focus of the fetus before the head came back to baseline position, the last frame of observation counted as the end of the movement. We excluded the partially captured head movements when calculating the duration of movements. We applied the same criteria used for the orofacial movements, when combining individual head movements into a single unit. In the article, when we refer to orofacial or head movements, we are speaking of movement units.

510 In young fetuses, where the body can be seen in addition to the head, other body parts were often  
511 moved with the head. We decided to focus on the head movements alone since the head region can  
512 be consistently observed with the orofacial region, even in the older fetuses where the body and the  
513 orofacial region cannot be imaged simultaneously. In addition, simulation experiments show that  
514 head and orofacial motor units are some of the last to differentiate [53].  
515

516  
517

#### 518 *Exemplars of fetal orofacial and head movements*

519

520 For the movement exemplars, the ultrasonography still images were generated by first using a  
521 custom-made MATLAB® program to split a chosen video clip into its component frames, and then  
522 selecting those frames which best exemplified the movement. The movement timeline plot was  
523 generated using onset and offset information of the movements (method described above) and  
524 plotted using MATLAB®. The amplitude of the movement did not have any bearing on the  
525 analysis.  
526

527

528

#### 529 *Information theory analysis - testing for increase in 'order' through gestation*

530

531 In our frame-by-frame analysis a movement counted as independent if the next movement began at  
532 least one frame after the completion of the movement in question. Independent movements were  
533 assigned as State 1 for orofacial movements and State 2 for head movements. If a head movement  
534 was initiated at least one frame after the beginning of the orofacial movement and before the end of  
535 the movement, the occurrence was assigned as State 3. If an orofacial movement was initiated at  
536 least one frame after beginning of the head movement and before the end of the movement, the  
537 occurrence was assigned as State 4. If orofacial and head movements were initiated on the same the  
538 frame, the occurrence was assigned as State 5 (simultaneous).  
539

540

541 Therefore, each movement was assigned to one of five states: independent orofacial movement  
542 (State1), independent head movement (State2), orofacial movement followed by an overlapping  
543 head movement (State3), head movement followed by an overlapping orofacial movement (State4),  
544 and orofacial plus head movements with synchronous onset (State5).  
545

546 The percentage of each movement state in one session was calculated by dividing the total number  
547 of movements of a particular state by total movement in the session (e.g., number of state1  
548 movements in session 1 divided by total number of movements in session 1). This was repeated for  
549 all sessions.

550

551 Shannon Entropy (H) for each session was calculated using the following formula:

552  $H(X) = -\sum_{i=1}^5 P(X = i) \log P(X = i)$  where the log is base 2 logarithm and i indicates the different  
553 states. To test if there was a change in entropy through gestation, we used the same curve fitting  
554 method used for movement and overlap rates.

555

556 Kullback-Leibler Divergence (KL) of Q from P was calculated using the following formula:



556  $KL(P|Q) = \sum_{i=1}^5 P(X=i) \log P(X=i) / Q(X=i)$  where  $i$  indicates different states and  $\log$  is in base  
557 2;  $P$  was the state distribution for the first testing session (gestational day 93);  $Q$  was the state  
558 distribution for every testing session (gestational days 93, 95....147). The same curve fitting method  
559 as above was used.

560

561 *Developmental change in the rates of orofacial and head movements*

562

563 The numbers of orofacial and head movements were counted for each session. Here too, roughly  
564 half of the total 64 sessions were coded by one experimenter, and the other half by another  
565 (alternate days done by the same person). Since the amount of time the face was visible differed  
566 between sessions, we calculated the rate of orofacial and head movements per hour. Movement rates  
567 were compiled across all pregnancies and polynomial curves (one for orofacial movements, one for  
568 head movements) were fitted to look at the trends across gestational time. To fit the curves, we first  
569 found the optimal degrees for polynomial fitting according to Akaike's information criterion (AIC)  
570 [72](MATLAB® polydeg). These degrees were then used in a polynomial curve fitting function  
571 (MATLAB® polyfit) to generate optimal fits for orofacial and head movement rates. To check if the  
572 observed trends held for individual pregnancies, we split the orofacial and head movement rates by  
573 pregnancy and used the optimal polynomial degrees calculated with the entire dataset to generate  
574 polynomial curves. The profiles of orofacial and head movement rates were found to be very robust,  
575 and the same profiles were seen when the analysis was repeated by looking at movement durations  
576 instead of movement numbers.

577

578

579 *Developmental change in the overlap of orofacial and head movements*

580

581 We counted the number of instances where orofacial and head movements occurred together and  
582 divided this number by the total number of orofacial and head movements (the same trend was  
583 found if the duration of overlap was used instead). The overlap calculation included instances where  
584 orofacial movements were followed by an overlapping head movement, instances where head  
585 movements were followed by an overlapping orofacial movement and instances where orofacial and  
586 head movements started simultaneously. To determine the trend across gestational time, we first  
587 found the optimal degrees for polynomial fitting according to Akaike's information criterion [72]  
588 (MATLAB® polydeg). This degree was then used in a polynomial curve fitting function  
589 (MATLAB® polyfit) to generate the curve. To check if the observed trends held at the pregnancy  
590 level, we split the overlap rates by pregnancy and used the optimal polynomial degrees calculated  
591 with the entire dataset to generate polynomial curves. To test for the significance of decrease in the  
592 overlap rate, we fitted a linear regression between the overlap rate and gestational date controlling  
593 for the pregnancies and tested for the nullity of the effect (ANOVA) of the gestational date.

594

595

596 *Permutation test*

597

598 The total number of orofacial and head movements was seen to decrease through gestation. To  
599 confirm that the observed decline in orofacial and head movement overlap was independent of the  
600 general decline in movement, a permutation test was performed. To do this, we took each session  
601 and independently reshuffled both the orofacial and head movements, keeping the duration and  
602 latency distributions intact. We then independently resampled with replacement, keeping the number

603 of durations and intervals the same as the original dataset. Overlap between orofacial and head  
604 movements was then calculated. This was done for every session. Average percentage of overlap was  
605 calculated for each gestational age (same procedure as that used in our original calculation). This  
606 procedure was repeated 1000 times and the upper 97.5 and lower 2.5 percentiles were computed to  
607 generate the confidence interval. To fit the curve, we followed the same procedure as with the  
608 observed data — we first found the optimal degrees for polynomial fitting according to AIC [72]  
609 (MATLAB® polydeg). This degree was then used in a polynomial curve fitting function  
610 (MATLAB® polyfit) to generate the curve. We expected this analysis to show the pattern of  
611 overlapping movements across gestation that would have resulted due to chance.

612  
613

#### 614 *Change in durations of orofacial and head movements through gestation*

615

616 Orofacial and head movement durations were calculated by subtracting a movement's offset time  
617 from its onset time. Movements that were marked as 'partially captured' during the initial scoring  
618 were excluded from this analysis. Median orofacial and head movement durations for each session  
619 were calculated and compiled across all sessions, for all pregnancies. To test if there was a change in  
620 the duration of orofacial and head movement through gestation, we used the same curve fitting  
621 method used for movement and overlap rates.

622

623

#### 624 *Experimental setup - infant vocalizations*

625

626 Audio and video recordings of infant vocalizations were made from P1 (~24 hours after birth) to  
627 P7. The infant was briefly separated from its caregivers and taken to a 2.5 m X 2.5 m room with  
628 walls covered in sound attenuating foam. For the audio recordings, the infant was placed on a layer  
629 of foam in a transfer cage. Once the subject was in place, the experimenter turned on a digital  
630 recorder (ZOOM H4n Handy Recorder) positioned directly in front of the testing cage at a distance  
631 of 0.76 m and left the room for a period of 5 minutes. For the video recordings one experimenter  
632 held the baby and a second experimenter acquired videos of the infant face using a hand-held  
633 SONY video recorder. This was done to observe facial movements during vocalizations. The frame  
634 rate of the infant videos matched that of the ultrasound videos — both were 30 frames per second.

635

636

#### 637 *Infant contact calls - audio and video recordings*

638

639 Audio recordings were processed using Adobe Audition software. The spectrograms of the audio  
640 signals were used to identify the infant contact call. In marking the calls, onset-offset gaps of 500ms  
641 or longer indicated separate calls, whereas gaps <500 ms indicated syllables from the same call [13].

642

643 Video analysis was done using Adobe Premiere Pro software, which enabled us to screen the videos  
644 frame-by-frame. The exact same frame-by-frame analysis used to score the fetal videos, was used for  
645 the infants. The first video frame where the jaws separated counted as the onset of orofacial  
646 movement and the first video frame where the jaws came back together was the offset of the  
647 movement. Here too, individual orofacial movements were considered to be part of the same  
648 movement unit if they were separated by <500ms (i.e., <15 frames) [13].

649

650 For our analysis, we used the P1 audio recordings and P1-7 video recordings. Marmoset infants  
651 frequently fall asleep on day 1 (just like any other newborn), especially when held. Therefore, to get a  
652 large enough sample size, we used the first week of video recording. The audio recordings on P1  
653 consisted of a total of 120 contact calls, the video recordings from P1-7 consisted of a total of 42  
654 contact calls. We combined both data to obtain enough samples.

655  
656 To compare the audio and video signals, we extracted audio from the video recordings, and  
657 measured call duration and syllable number separately in each modality. We then measured the  
658 difference in duration and syllable number in the two signals. In case of duration, the video  
659 measures were consistently longer than the audio. To compensate for the duration difference, we  
660 calculated the difference between the video and audio duration. The median duration discrepancy  
661 was added to the audio duration of all contact calls to adjust for the discrepancy. The number of  
662 syllables remained the same across the audio and video measures; no adjustment was required.

663  
664

#### 665 *Infant contact call exemplar*

666  
667 The image panels were generated by first using a custom-made MATLAB® program to split the  
668 chosen infant contact call video clip into its component frames and then selecting those frames  
669 which best exemplified the movement.

670  
671 The spectrogram was produced by extracting the contact call audio signal from the chosen clip and  
672 reducing background noise using the Adobe Audition software.

673  
674 The temporal profiles of the accompanying orofacial movements were generated using a custom  
675 MATLAB® program which allowed us to go through the movement clip frame-by-frame and mark  
676 the positions of the lower and upper jaws by selecting a central point on each jaw. (When there is no  
677 orofacial movement, the distance between the two points is 0, while a nonzero value indicates that  
678 the mouth is open.) The profiles were then z-scored to remove amplitude information. The  
679 generated profiles were smoothed using the cubic smoothing spline (csaps) function in MATLAB®  
680 (smoothing parameter = .999).

681  
682

#### 683 *Fetal contact call precursor exemplar*

684  
685 Temporal profiles of a set of fetal orofacial movements that were similar in duration and syllable  
686 number to the infant contact call were generated using the same custom MATLAB® program and  
687 procedure used for generating infant contact call profiles. A profile that best matched the infant  
688 contact call profile was selected as the exemplar.

689  
690

#### 691 692 *Matching temporal profiles of fetal orofacial movements to that of week 1 infant contact calls.*

693  
694 Videos of P1-7 infants producing contact calls were compiled (n=25), and a custom MATLAB®  
695 program was used to do a frame-by-frame tracking of the upper and lower jaw positions to create  
696 temporal profiles for the contact calls. These calls consisted of 1-7 syllables (1 syllable, n=2; 2  
697 syllables, n=3; 3 syllables, n=5; 4 syllables, n=5; 5 syllables, n=2; 6 syllables, n=4; 7 syllable, n=4).

698 The profiles generated from calls of the same syllable number were averaged using the DTW  
699 Barycenter Averaging (DBA) MATLAB® routine [73]. The average profiles were then smoothed  
700 using MATLAB® spline smoothing function csaps (smoothing parameter=.1), to generate the  
701 templates of infant contact calls (n=7).

702  
703 A custom computer vision program was developed in C++ to track fetal orofacial movements in an  
704 automated fashion. The program was set up with a MATLAB® interface. Through this interface we  
705 loaded individual movement clips, and marked the head, nose and orofacial regions of the fetus on  
706 the first frame of the clip. The program then went through the rest of the clip frame-by-frame and  
707 generated the temporal profile of the orofacial movement.

708  
709 Fetuses move significantly in the womb, especially in the mid stages of gestation. To make sure we  
710 only included accurate tracings of the movements, we scored the temporal profiles of all the  
711 movements (n=1977). All the partially captured movements were excluded, and only traced profiles  
712 with highest score—those that matched the manually calculated syllable number features of the  
713 movement—were included in the rest of the analysis (n=460).

714  
715 The temporal profiles were then smoothed using the Savitzky-Golay Smoothing Filter. The tuning  
716 parameters we used were: polynomial degree = 3 and sliding window size = 9. The tuning  
717 parameters were selected based on what retained the structure of the tracings and could be used for  
718 the majority of movements (n=414).

719

720

721 *Dynamic Time Warping (DTW) analysis*

722

723 To measure the similarity between the two temporal sequences (fetal orofacial movement and infant  
724 contact call template) with possibly different time lengths, we used the continuous DTW algorithm  
725 using a linear interpolation model. We used the DTW Python module generated by Pierre Rouanet  
726 (version 1.3.3). The cost of the DTW was used as a measure of similarity between two signals.  
727 Smaller values of DTW cost indicate larger similarity between the signals. Each fetal movement  
728 profile was compared with all seven infant contact call templates, and a DTW score generated. The  
729 lowest DTW score was retained. DTW scores from all the movements in a single session were  
730 compiled and the median DTW score calculated. Median DTW scores were generated for all  
731 sessions (n=57). The statistical computing was done using R. We fitted a multiple linear regression  
732 model to predict similarity to the infant contact call temporal profile (DTW score) based on  
733 gestational age and pregnancy.

734

735

736

737 *Gestational change in the duration and 'syllable numbers' of fetal orofacial movements*

738

739 Orofacial movement durations were calculated by subtracting a movement's offset time from its  
740 onset time. Syllable numbers were calculated by counting the number of mouth movements in each  
741 movement unit. Partially captured movements were not included in this analysis. Median, 75th  
742 percentile, and 25th percentile values for duration and number of individual movements in a  
743 movement unit (equivalent of syllable number) were calculated for each session. The developmental  
744 trajectories of duration and syllable number were generated by first calculating the optimal degree  
745 for polynomial fitting according to AIC [72] (MATLAB® polydeg), for the median values. The same

746 polynomial fitting degree was used for the other two curves. The contour plots for median durations  
747 and syllable numbers were made using MATLAB routine ‘Kernel density estimation’ by Zdravko  
748 Botev (MATLAB file exchange #17204).

749  
750

751 *Duration and syllable number features of the infant contact calls at P1*

752

753 The boxplots for P1 infant contact call duration and syllable number distributions were done using  
754 the ‘Box and whiskers plot (without statistics toolbox)’ by Jonathan C.Lanseby (MATLAB file  
755 exchange #42470). These plots indicate the median, 75th percentile, and 25th percentile values.

756

757

758 *Criteria for defining contact calls based on duration and syllable number features*

759

760 To be able to match the durations and syllable numbers of fetal orofacial movements to the infant  
761 contact call, we had to obtain criteria based on orofacial movement that allowed us to select contact  
762 calls with high probability. Following is the method we used for setting the bounds. (1) For every  
763 call duration (500ms - 6.5s) or syllable number (1-9), we calculated the likelihood that the call is a  
764 contact call based on our infant contact call samples [e.g., if the duration of a call is 6s the likelihood  
765 that it is a contact call is 100%] (2) We set the duration and syllable number range for contact calls to  
766 ensure maximum possible separation from other call types. The duration criteria we arrived at for  
767 the infant contact call is 3.69-6.5s. The syllable number criteria we arrived at is 5-9. When both the  
768 duration and syllable number criteria are applied, the contact call likelihood is 97%.

769

770

771 *Matching fetal orofacial movements to contact call criteria*

772

773 The percentage of fetal orofacial movements matching the contact call duration+syllable number  
774 criteria (‘percent match contact call dur-syll’) was calculated for each session. The statistics were  
775 calculated using R. We fitted a multiple linear regression model to predict percentage match of  
776 contact call dur-syll based on gestational age and pregnancy. ANOVA test was used to test for the  
777 significance of the effect of gestational age.

778

779

780 *Criteria for defining twitter calls based on duration and syllable number features*

781

782 The criteria for defining twitter calls was obtained using the same method as that used for the  
783 contact calls. The duration criteria we arrived at for the infant twitter call is 0.96-1.5s. The syllable  
784 number criteria we arrived at is 2-3. A call within this duration and syllable number criteria has an  
785 86% likelihood of being a twitter call and a 2% likelihood of being a contact call.

786

787

788 *Matching fetal orofacial movements to twitter call criteria*

789

790 The percentage of fetal orofacial movements matching the twitter duration+syllable number criteria  
791 (‘percent match twitter dur-syll’) was calculated for each session. The statistics were calculated using  
792 R. We fitted a multiple linear regression model to predict the percentage match of twitter dur-syll

793 based on gestational age and pregnancy. ANOVA test was used to test for the significance of the  
794 effect of gestational age.

795

796

797 *Criteria for defining infant licks based on duration and syllable number features*

798

799 Video analysis was done using Adobe Premiere Pro software, which enabled us to screen the videos  
800 frame-by-frame. The exact same frame-by-frame analysis used to score the fetal videos and infant  
801 vocalizations was used for the licking movements. The first video frame where the jaws separated  
802 counted as the onset of orofacial movement and the first video frame where the jaws came back  
803 together was the offset of the movement. Here too, individual orofacial movements were considered  
804 to be part of the same movement unit if they were separated by <500ms (i.e., <15 frames) [13].  
805 Orofacial movement durations were calculated by subtracting a movement's offset time from its  
806 onset time. Syllable numbers were calculated by counting the number of mouth movements in each  
807 movement unit.

808

809 The licks of P1-7 infants (n=37) tend to be short with few syllables, or very long with a high number  
810 of syllables. Therefore, to capture the maximum possible infant licks, we established a separate  
811 criteria for each of these types of licks. Short licks: duration range 0.33–3.77s and syllable number  
812 range 1-6. Long licks: duration range 6.67–14.29s and syllable number range 10-16. Following is the  
813 method we used for determining the bounds: 0.33 and 14.29 are the 2.5 and 97.5 percentiles of the  
814 lick durations; while 1 and 16 are the 2.5 and 97.5 percentiles of the lick syllable number. There is  
815 only 1 movement of duration between the 3.77 and 6.67s (therefore our duration criteria only  
816 filtered out 0.03% of the total movements). There are 3 movements of syllable number between 6  
817 and 10 (therefore our syllable number criteria only filtered out 0.08% of the movements). Together  
818 our criteria capture 83.78% of the infant licks (and 0% of the infant contact calls).

819

820

821 *Matching fetal orofacial movements to lick call criteria*

822

823 The percentage of fetal orofacial movements matching the lick duration+syllable number criteria  
824 ('percent match lick dur-syll') was calculated for each session. The statistics were calculated using R.  
825 We fitted a multiple linear regression model to predict the percentage match of lick dur-syll based on  
826 gestational age and pregnancy. ANOVA test was used to test for the significance of the effect of  
827 gestational age.

828

829

830

- 831  
832  
833 1. Hopkins B, Von Wulfften Palthe T. The development of the crying state during early  
834 infancy. *Developmental Psychobiology: The Journal of the International Society for*  
835 *Developmental Psychobiology*. 1987;20(2):165-75.  
836 2. Wolff PH. The natural history of crying and other vocalizations in early infancy.  
837 *Determinants of infant behavior*. 1969:81-111.  
838 3. Chandrasekaran C, Trubanova A, Stillitano S, Caplier A, Ghazanfar AA. The natural  
839 statistics of audiovisual speech. *PLoS computational biology*. 2009;5(7):e1000436.  
840 4. Risueno-Segovia C, Hage SR. Theta synchronization of phonatory and articulatory  
841 systems in marmoset monkey vocal production. *Current Biology*. 2020;30(21):4276-83. e3.  
842 5. Lameira AR, Hardus ME, Bartlett AM, Shumaker RW, Wich SA, Menken SB. Speech-  
843 like rhythm in a voiced and voiceless orangutan call. *PLoS one*. 2015;10(1):e116136.  
844 6. Ghazanfar AA, Takahashi DY. The evolution of speech: vision, rhythm, cooperation.  
845 *Trends in cognitive sciences*. 2014;18(10):543-53.  
846 7. Bergman TJ, Beehner JC, Painter MC, Gustison ML. The speech-like properties of  
847 nonhuman primate vocalizations. *Animal Behaviour*. 2019;151:229-37.  
848 8. Goldstein MH, Schwade JA. Social feedback to infants' babbling facilitates rapid  
849 phonological learning. *Psychological science*. 2008;19(5):515-23.  
850 9. Warlaumont AS, Richards JA, Gilkerson J, Oller DK. A social feedback loop for speech  
851 development and its reduction in autism. *Psychological science*. 2014;25(7):1314-24.  
852 10. Mampe B, Friederici AD, Christophe A, Wermke K. Newborns' cry melody is shaped by  
853 their native language. *Current biology*. 2009;19(23):1994-7.  
854 11. Gingras JL, Mitchell EA, Grattan KE. Fetal homologue of infant crying. *Arch Dis Child*  
855 *Fetal Neonatal Ed*. 2005;90(5):F415-8. doi: 10.1136/adc.2004.062257. PubMed Central PMCID:  
856 PMC1721928.  
857 12. Takahashi DY, Fenley AR, Teramoto Y, Narayanan DZ, Borjon JI, Holmes P, et al. The  
858 developmental dynamics of marmoset monkey vocal production. *Science*. 2015;349(6249):734-  
859 8. doi: 10.1126/science.aab1058.  
860 13. Zhang YS, Ghazanfar AA. Perinatally Influenced Autonomic System Fluctuations Drive  
861 Infant Vocal Sequences. *Curr Biol*. 2016;26(10):1249-60. doi: 10.1016/j.cub.2016.03.023.  
862 14. Takahashi DY, Liao DA, Ghazanfar AA. Vocal learning via social reinforcement by  
863 infant marmoset monkeys. *Current Biology*. 2017;27(12):1844-52. e6.  
864 15. Gultekin YB, Hage SR. Limiting parental feedback disrupts vocal development in  
865 marmoset monkeys. *Nature Communications*. 2017;8:14046. doi: 10.1038/ncomms14046.  
866 16. Gultekin YB, Hage SR. Limiting parental interaction during vocal development affects  
867 acoustic call structure in marmoset monkeys. *Sci Adv*. 2018;4(4):eaar4012. doi:  
868 10.1126/sciadv.aar4012. PubMed Central PMCID: PMC5895450.  
869 17. de Castro Leão A, Neto ADD, de Sousa MBC. New developmental stages for common  
870 marmosets (*Callithrix jacchus*) using mass and age variables obtained by K-means algorithm and  
871 self-organizing maps (SOM). *Computers in biology and medicine*. 2009;39(10):853-9.  
872 18. Harris RA, Tardif SD, Vinar T, Wildman DE, Rutherford JN, Rogers J, et al.  
873 Evolutionary genetics and implications of small size and twinning in callitrichine primates. *Proc*  
874 *Natl Acad Sci U S A*. 2014;111:1467-72.

- 875 19. Fagard J, Esseily R, Jacquy L, O'Regan K, Somogyi E. Fetal Origin of Sensorimotor  
876 Behavior. *Front Neurobot.* 2018;12:23. doi: 10.3389/fnbot.2018.00023. PubMed Central  
877 PMCID: PMC5974044.
- 878 20. Kurjak A, Stanojevic M, Andonotopo W, Salihagic-Kadic A, Carrera JM, Azumendi G.  
879 Behavioral pattern continuity from prenatal to postnatal life--a study by four-dimensional (4D)  
880 ultrasonography. *J Perinat Med.* 2004;32(4):346-53. doi: 10.1515/JPM.2004.065.
- 881 21. Bekoff A, Lau B. Interlimb coordination in 20-day-old rat fetuses. *J Exp Zool.*  
882 1980;214(2):173-5. doi: 10.1002/jez.1402140207.
- 883 22. Brumley MR, Robinson SR. Experience in the perinatal development of action systems.  
884 *Oxford handbook of developmental.* 2010.
- 885 23. Smotherman WP, Robinson SR. The uterus as environment. *The ecology of fetal*  
886 *behaviour. Handbook of Behavioral Neurobiology.* 1988;9:149-96.
- 887 24. Bradley NS. Transformations in embryonic motility in chick: kinematic correlates of type  
888 I and II motility at E9 and E12. *J Neurophysiol.* 1999;81(4):1486-94.
- 889 25. Hamburger V. Some Aspects of the Embryology of Behavior. *Q Rev Biol.*  
890 1963;38(4):342-. doi: 10.1086/403941.
- 891 26. Kuo ZY. Ontogeny of embryonic behavior in aves. IV. The influence of embryonic  
892 movements upon the behavior after hatching. *J Comp Psychol.* 1932;14(1):109-22. doi:  
893 10.1037/h0071451.
- 894 27. Provine RR. Development of between-limb movement synchronization in the chick  
895 embryo. *Dev Psychobiol.* 1980;13(2):151-63. doi: 10.1002/dev.420130207.
- 896 28. Keogh E, Ratanamahatana CA. Exact indexing of dynamic time warping. *Knowledge and*  
897 *information systems.* 2005;7(3):358-86.
- 898 29. Ghazanfar AA, Liao DA. Constraints and flexibility during vocal development: Insights  
899 from marmoset monkeys. *Curr Opin Behav Sci.* 2018;21:27-32. doi:  
900 10.1016/j.cobeha.2017.11.015. PubMed Central PMCID: PMC5982589.
- 901 30. Teramoto Y, Takahashi DY, Holmes P, Ghazanfar AA. Vocal development in a  
902 Waddington landscape. *Elife.* 2017;6. doi: 10.7554/eLife.20782. PubMed Central PMCID:  
903 PMC5310845.
- 904 31. Thelen E, Krasnegor NA, Rumbaugh DM, Schiefelbusch RL, Studdert-Kennedy M.  
905 Biological and behavioral determinants of language development. *Biological and behavioral*  
906 *determinants of language development* Hilldale, New Jersey: Lawrence Erlbaum Associates.  
907 1991.
- 908 32. Huang J, Cheng X, Zhang S, Chang L, Li X, Liang Z, et al. Having Infants in the Family  
909 Group Promotes Altruistic Behavior of Marmoset Monkeys. *Current Biology.* 2020;30(20):4047-  
910 55. e3.
- 911 33. Takahashi DY, Fenley AR, Ghazanfar AA. Early development of turn-taking with  
912 parents shapes vocal acoustics in infant marmoset monkeys. *Philos Trans R Soc Lond B Biol*  
913 *Sci.* 2016;371(1693). doi: 10.1098/rstb.2015.0370. PubMed Central PMCID:  
914 PMC4843608.
- 915 34. Barron, Donald BY, H. The functional development of some mammalian neuromuscular  
916 mechanisms. *Biol Rev Camb Philos Soc.* 1941;16(1):1-33.
- 917 35. Coghill GE. *Anatomy and the problem of behavior:* MacMillan Co., New York; 1929  
918 1929. 113- p.
- 919 36. Oppenheim RW. The ontogeny of behavior in the chick embryo. *Advances in the Study*  
920 *of Behavior Vol 5:* Academic Press; 1974. p. 133-72.



- 921 37. Smotherman WP, Robinson R. Dimensions of fetal investigation. The Teleford Press,  
922 INC.; 1988. p. 19-34.
- 923 38. Granmo M, Petersson P, Schouenborg J. Action-based body maps in the spinal cord  
924 emerge from a transitory floating organization. *J Neurosci*. 2008;28(21):5494-503. doi:  
925 10.1523/JNEUROSCI.0651-08.2008.
- 926 39. Barlow SM, Radder JPL, Radder ME, Radder AK. Central pattern generators for  
927 orofacial movements and speech. *Handbook of behavioral neuroscience*. 2010;19:351-69.
- 928 40. Moore JD, Kleinfeld D, Wang F. How the brainstem controls orofacial behaviors  
929 comprised of rhythmic actions. *Trends in neurosciences*. 2014;37(7):370-80.
- 930 41. Bekoff A. Spontaneous embryonic motility: an enduring legacy. *Int J Dev Neurosci*.  
931 2001;19(2):155-60. doi: 10.1016/S0736-5748(00)00089-7.
- 932 42. Gottlieb G. Conceptions of prenatal development: Behavioral embryology. *Psychol Rev*.  
933 1976;83(3):215-34. doi: 10.1037//0033-295X.83.3.215.
- 934 43. Roodenburg PJ, Wladimiroff JW, van Es A, Prectl HFR. Classification and quantitative  
935 aspects of fetal movements during the second half of normal pregnancy. *Early Hum Dev*.  
936 1991;25(1):19-35. doi: 10.1016/0378-3782(91)90203-F.
- 937 44. Decker JD, Hamburger V. The influence of different brain regions on periodic motility of  
938 the chick embryo. *J Exp Zool*. 1967;165(3):371-83. doi: 10.1002/jez.1401650306.
- 939 45. Marques HG, Bharadwaj A, Iida F. From spontaneous motor activity to coordinated  
940 behaviour: a developmental model. *PLoS Comput Biol*. 2014;10(7):e1003653.
- 941 46. Kleven GA, Lane MS, Robinson SR. Development of interlimb movement synchrony in  
942 the rat fetus. *Behav Neurosci*. 2004;118:835-44. doi: 10.1037/0735-7044.118.4.835.
- 943 47. Robinson SR, Kleven GA. Learning to move before birth. 2005. p. 131-75.
- 944 48. Bradley NS, Sebelski C. Ankle restraint modifies motility at E12 in chick embryos. *J*  
945 *Neurophysiol*. 2000;83(1):431-40. doi: 10.1152/jn.2000.83.1.431.
- 946 49. Sharp AA, Ma E, Bekoff A. Developmental changes in leg coordination of the chick at  
947 embryonic days 9, 11, and 13: uncoupling of ankle movements. *J Neurophysiol*.  
948 1999;82(5):2406-14. doi: 10.1152/jn.1999.82.5.2406.
- 949 50. Kuniyoshi Y. Fusing autonomy and sociality via embodied emergence and development  
950 of behaviour and cognition from fetal period. *Philos Trans R Soc Lond B Biol Sci*.  
951 2019;374(1771):20180031. doi: 10.1098/rstb.2018.0031.
- 952 51. Mori H, Kuniyoshi Y. A human fetus development simulation: Self-organization of  
953 behaviors through tactile sensation. 2010 IEEE 9th International Conference on Development  
954 and Learning. 2010:82-7. doi: 10.1109/DEVLRN.2010.5578860.
- 955 52. Yamada Y, Mori H, Kuniyoshi Y. A Fetus and Infant Developmental Scenario: Self-  
956 organization of Goal-directed Behaviors Based on Sensory Constraints. *Proceedings of the Tenth*  
957 *International Conference on Epigenetic Robotics: Modeling Cognitive Development in Robotic*  
958 *Systems*. 2010.
- 959 53. Yamada Y, Kuniyoshi Y. Embodiment guides motor and spinal circuit development in  
960 vertebrate embryo and fetus. 2012 IEEE International Conference on Development and Learning  
961 and Epigenetic Robotics (ICDL). 2012:1-6. doi: 10.1109/DevLrn.2012.6400578.
- 962 54. Yamada Y, Kanazawa H, Iwasaki S, Tsukahara Y, Iwata O, Yamada S, et al. An  
963 Embodied Brain Model of the Human Foetus. *Sci Rep*. 2016;6:27893. doi: 10.1038/srep27893.  
964 PubMed Central PMCID: PMC4908381.

- 965 55. Mori H. Is the developmental order of fetal behaviors self-organized in an uterine  
966 environment? 2012 IEEE International Conference on Development and Learning and  
967 Epigenetic Robotics (ICDL). 2012:1-2.
- 968 56. Dooley JC, Blumberg MS. Developmental'awakening'of primary motor cortex to the  
969 sensory consequences of movement. *elife*. 2018;7:e41841.
- 970 57. Dooley JC, Glanz RM, Sokoloff G, Blumberg MS. Self-generated whisker movements  
971 drive state-dependent sensory input to developing barrel cortex. *Current Biology*. 2020.
- 972 58. Blumberg MS, Coleman CM, Gerth AI, McMurray B. Spatiotemporal structure of REM  
973 sleep twitching reveals developmental origins of motor synergies. *Current Biology*.  
974 2013;23(21):2100-9.
- 975 59. Blumberg MS, Marques HG, Iida F. Twitching in Sensorimotor Development from  
976 Sleeping Rats to Robots. *Curr Biol*. 2013;23(12):R532-R7.
- 977 60. Newman JD. The primate isolation call: a comparison with precocial birds and non-  
978 primate mammals. *Comparative vertebrate cognition*: Springer; 2004. p. 171-87.
- 979 61. Oller DK. The emergence of the speech capacity: Psychology Press; 2000 2000.
- 980 62. Ross CN, Power ML, Artavia JM, Tardif SD. Relation of food intake behaviors and  
981 obesity development in young common marmoset monkeys. *Obesity*. 2013;21:1891-9.
- 982 63. Adolph KE, Robinson SR, Young JW, Gill-Alvarez F. What is the shape of  
983 developmental change? *Psychological review*. 2008;115(3):527.
- 984 64. Burkart JM, Allon O, Amici F, Fichtel C, Finkenwirth C, Heschl A, et al. The  
985 evolutionary origin of human hyper-cooperation. *Nature communications*. 2014;5(1):1-9.
- 986 65. Ghazanfar AA, Kelly LM, Takahashi DY, Winters S, Terrett R, Higham JP.  
987 Domestication phenotype linked to vocal behavior in marmoset monkeys. *Current Biology*.  
988 2020;30(24):5026-32. e3.
- 989 66. Choi JY, Takahashi DY, Ghazanfar AA. Cooperative vocal control in marmoset monkeys  
990 via vocal feedback. *Journal of neurophysiology*. 2015;114(1):274-83.
- 991 67. Takahashi DY, Narayanan DZ, Ghazanfar AA. Coupled Oscillator Dynamics of Vocal  
992 Turn-Taking in Monkeys. *Current Biology*. 2013 2013.
- 993 68. Zürcher Y, Willems EP, Burkart JM. Are dialects socially learned in marmoset monkeys?  
994 Evidence from translocation experiments. *PloS one*. 2019;14(10):e0222486.
- 995 69. Liao DA, Zhang YS, Cai LX, Ghazanfar AA. Internal states and extrinsic factors both  
996 determine monkey vocal production. *Proceedings of the National Academy of Sciences*.  
997 2018;115(15):3978-83.
- 998 70. Varella TT, Ghazanfar AA. Cooperative care and the evolution of the prelinguistic vocal  
999 learning. *Developmental psychobiology*. 2021;In press.
- 1000 71. Jaquish CE, Toal RL, Tardif SD, Carson RL. Use of ultrasound to monitor prenatal  
1001 growth and development in the common marmoset (*Callithrix jacchus*). *American journal of*  
1002 *primatology*. 1995;36(4):259-75.
- 1003 72. Akaike H. A new look at the statistical-model identification.  
1004 *Current/Contents/Engineering Technology & Applied Sciences*. 1981;(51):22-.
- 1005 73. Petitjean F, Ketterlin A, Gançarski P. A global averaging method for dynamic time  
1006 warping, with applications to clustering. *Pattern Recognit*. 2011;44(3):678-93. doi:  
1007 10.1016/j.patcog.2010.09.013.