SUPPORTING INFORMATION (SI) APPENDIX

1 GPDC values by EEG frequency band

1.1 Experiment 1 : Video

		Across Individuals								Within Individuals			
		Adult -> Infant				Infant -> Adult				Infant -> Infant		Adult -> Adult	
		L		R		L		R		L	R	L	R
		L	R	L	R	L	R	L	R	R	L	R	L
Theta *(3-6 Hz)*	Direct	**0.076**	**0.082**	**0.087**	**0.090**	**0.068**	**0.067**	**0.071**	**0.070**	**0.135**	**0.134**	**0.065**	**0.078**
		0.011	*0.016*	*0.014*	*0.008*	*0.009*	*0.008*	*0.008*	*0.009*	*0.014*	*0.018*	*0.006*	*0.006*
	Indirect	**0.076**	**0.082**	**0.081**	**0.084**	**0.069**	**0.075**	**0.067**	**0.068**	**0.137**	**0.135**	**0.066**	**0.065**
		0.011	*0.015*	*0.013*	*0.007*	*0.007*	*0.011*	*0.007*	*0.007*	*0.013*	*0.017*	*0.007*	*0.010*
	Direct-Oblique	**0.079**	**0.084**	**0.090**	**0.086**	**0.067**	**0.068**	**0.064**	**0.067**	**0.141**	**0.138**	**0.063**	**0.078**
		0.013	*0.012*	*0.012*	*0.012*	*0.006*	*0.007*	*0.007*	*0.008*	*0.016*	*0.014*	*0.006*	*0.009*
	Surrogate	**0.065**	**0.068**	**0.074**	**0.077**	**0.068**	**0.069**	**0.068**	**0.069**	**0.118**	**0.114**	**0.057**	**0.062**
		0.004	*0.005*	*0.006*	*0.008*	*0.003*	*0.003*	*0.003*	*0.003*	*0.011*	*0.010*	*0.001*	*0.001*
Alpha *(6-9 Hz)*	Direct	**0.078**	**0.088**	**0.091**	**0.094**	**0.043**	**0.045**	**0.045**	**0.045**	**0.138**	**0.136**	**0.044**	**0.055**
		0.012	*0.015*	*0.016*	*0.012*	*0.004*	*0.005*	*0.007*	*0.005*	*0.019*	*0.019*	*0.005*	*0.005*
	Indirect	**0.080**	**0.087**	**0.085**	**0.091**	**0.043**	**0.048**	**0.042**	**0.045**	**0.140**	**0.136**	**0.043**	**0.043**
		0.011	*0.016*	*0.017*	*0.010*	*0.005*	*0.006*	*0.003*	*0.006*	*0.018*	*0.019*	*0.004*	*0.006*
	Direct-Oblique	**0.084**	**0.088**	**0.096**	**0.095**	**0.042**	**0.045**	**0.041**	**0.045**	**0.138**	**0.135**	**0.043**	**0.048**
		0.011	*0.013*	*0.011*	*0.017*	*0.004*	*0.004*	*0.005*	*0.006*	*0.017*	*0.019*	*0.005*	*0.005*
	Surrogate	**0.071**	**0.075**	**0.081**	**0.085**	**0.042**	**0.044**	**0.042**	**0.044**	**0.119**	**0.115**	**0.037**	**0.039**
		0.006	*0.007*	*0.008*	*0.010*	*0.002*	*0.002*	*0.002*	*0.002*	*0.015*	*0.014*	*0.001*	*0.001*

Shaded connections were not significantly above surrogate threshold (BH FDR-corrected at p<.05). See Section 10 for surrogate analysis.

*Table S1 – Experiment 1 GPDC values by EEG frequency band (**mean** in bold, SD in italics)*

1.2 Experiment 2 : Live

		Across Individuals								Within Individuals			
		Adult -> Infant				Infant -> Adult				Infant -> Infant		Adult -> Adult	
		L		R		L		R		L	R	L	R
		L	R	L	R	L	R	L	R	R	L	R	L
Theta *(3-6 Hz)*	Direct	**0.072**	**0.071**	**0.070**	**0.072**	**0.068**	**0.060**	**0.067**	**0.068**	**0.100**	**0.110**	**0.075**	**0.077**
		0.025	*0.026*	*0.023*	*0.028*	*0.019*	*0.009*	*0.019*	*0.017*	*0.032*	*0.038*	*0.021*	*0.020*
	Indirect	**0.072**	**0.064**	**0.068**	**0.066**	**0.061**	**0.064**	**0.066**	**0.061**	**0.100**	**0.106**	**0.073**	**0.080**
		0.022	*0.017*	*0.019*	*0.019*	*0.017*	*0.012*	*0.015*	*0.011*	*0.025*	*0.033*	*0.018*	*0.022*
	Surrogate	**0.058**	**0.058**	**0.059**	**0.059**	**0.059**	**0.058**	**0.059**	**0.060**	**0.074**	**0.075**	**0.056**	**0.057**
		0.012	*0.010*	*0.013*	*0.012*	*0.010*	*0.008*	*0.010*	*0.007*	*0.012*	*0.013*	*0.009*	*0.009*
Alpha *(6-9 Hz)*	Direct	**0.075**	**0.077**	**0.072**	**0.080**	**0.042**	**0.037**	**0.041**	**0.041**	**0.101**	**0.104**	**0.049**	**0.048**
		0.022	*0.029*	*0.027*	*0.034*	*0.010*	*0.005*	*0.011*	*0.008*	*0.038*	*0.038*	*0.014*	*0.012*
	Indirect	**0.072**	**0.066**	**0.069**	**0.066**	**0.037**	**0.041**	**0.039**	**0.038**	**0.100**	**0.107**	**0.048**	**0.051**
		0.024	*0.027*	*0.025*	*0.023*	*0.010*	*0.010*	*0.007*	*0.008*	*0.035*	*0.039*	*0.014*	*0.014*
	Surrogate	**0.061**	**0.062**	**0.061**	**0.062**	**0.035**	**0.035**	**0.035**	**0.036**	**0.073**	**0.073**	**0.034**	**0.034**
		0.019	*0.018*	*0.020*	*0.021*	*0.005*	*0.005*	*0.005*	*0.004*	*0.024*	*0.026*	*0.005*	*0.004*

Shaded connections were not significantly above surrogate threshold (BH FDR-corrected at p<.05). See Section 10 for surrogate analysis.

*Table S2 – Experiment 2 GPDC values by EEG frequency band (**mean** in bold, SD in italics)*

2 Infant vocalisation analysis and correlations with neural connectivity

Infants' vocalisations were manually coded from videos recorded during the experimental session according to Oller's [1] infraphonological acoustic classification system. This coding scheme incorporates acoustic features (such as fundamental frequency and formant transitions) with qualitative descriptors (e.g. phonetic categories) to distinguish between four categories of vocalisations : quasi-resonant vowel nuclei, fully-resonant vowel nuclei, marginal syllables and canonical syllables. The infants in both studies (median age of 8/8.5 months) were expected to produce all four categories of vocalisations. The total mean number and (utterance) duration of infants' vocalisations in each experiment and gaze condition are shown in Table S3.

	Gaze Condition	Mean number per infant	Mean duration per infant (s)
Expt 1	Direct	**8.22** *(2.43)*	**0.69** *(.10)*
	Indirect	**7.44** *(1.80)*	**0.82** *(.15)*
	Direct-Oblique	**7.11** *(1.69)*	**0.70** *(.07)*
Expt 2	Direct	**6.32** *(1.11)*	**0.80** *(.10)*
	Indirect	**5.00** *(1.20)*	**0.85** *(.08)*

Table S3. Mean number and duration of infants' vocalisations in each experiment and gaze condition. Means are shown in bold, standard errors are shown in italics.

As discussed in the main manuscript, the total number of utterances increased significantly for Direct relative to Indirect gaze in Expt 2 (live), but not for Expt 1 (video) where interactions were uni-directional (the infant could not influence the adult). To assess whether the social interaction context significantly moderated the effect of gaze on infants' vocalisations, we computed the mean difference between infants' number of vocalisations under Direct versus Indirect gaze, predicting that this difference would be larger for Expt 2 (live) than for Expt 1 (video). A normalized (i.e. Direct minus Indirect) index was used rather than raw values to account for differences in the baseline number of vocalisations between experiments. As there were two Direct gaze conditions in Experiment 1 (Direct and Direct-Oblique), we used the average number of vocalisations across these two conditions. An ANOVA was then conducted on the normalized vocalisation index, taking Experiment as the between-subjects factor, and controlling for infants' looking time and age. This analysis revealed a trend toward a significant difference between experiments ($F(1,23)=2.17$, $p=.077$, one-tailed), with a larger benefit of Direct gaze for vocalisations in the live interaction context (Expt 1) than for video (Expt 2), as predicted.

This live benefit is reminiscent of Kuhl et al's study [2] where infants showed phoneme learning from live speakers but not from video DVDs of the same speakers. Similarly, Goldstein & Schwade [3] found that only infants who received live *contingent* feedback from their mothers showed re-structuring of their babbling patterns. Consistent with these studies, here, infants produced more vocalisations only when Direct gaze was offered in a live contingent context.

2.1 Vocalisations by category

A breakdown of the mean number of infant vocalisations by category of complexity is provided in Figure S1. For each experiment, a Repeated Measures ANOVA was conducted to assess whether the complexity of vocalisations differed across gaze conditions, taking Complexity (4 levels) and Gaze (3 or 2 levels) as within-subjects factors. For Experiment 1, there was no main effect of Gaze ($F(2,32) = .29$, $p=.75$, $\eta^2p = .02$) and no interaction between Gaze and

vocalisation Complexity ($F(6,96) = .39$, $p = .88$, $\eta^2p = .02$). However, there was a significant main effect of Complexity ($F(3,48) = 8.94$, $p < .001$, $\eta^2p = .36$. Significantly more quasi-resonant and fully-resonant nuclei were produced than marginal and canonical syllables, but there was no difference within these sub-categories.

For Experiment 2, there was a significant main effect of Gaze ($F(1,18) = 5.80$, $p < .05$, $\eta^2p = .24$) but no interaction between Gaze and vocalisation Complexity ($F(3,54) = 1.67$, $p = .18$, $\eta^2p = .09$). However, there was again a significant main effect of Complexity ($F(3,54) = 8.20$, $p < .001$, $\eta^2p = .31$. As for Expt 1, significantly more quasi-resonant and fully-resonant nuclei were produced than marginal and canonical syllables, but there was again no difference within these sub-categories. Therefore, these results indicate that the adult speaker's gaze did not change the complexity of infants' utterances.

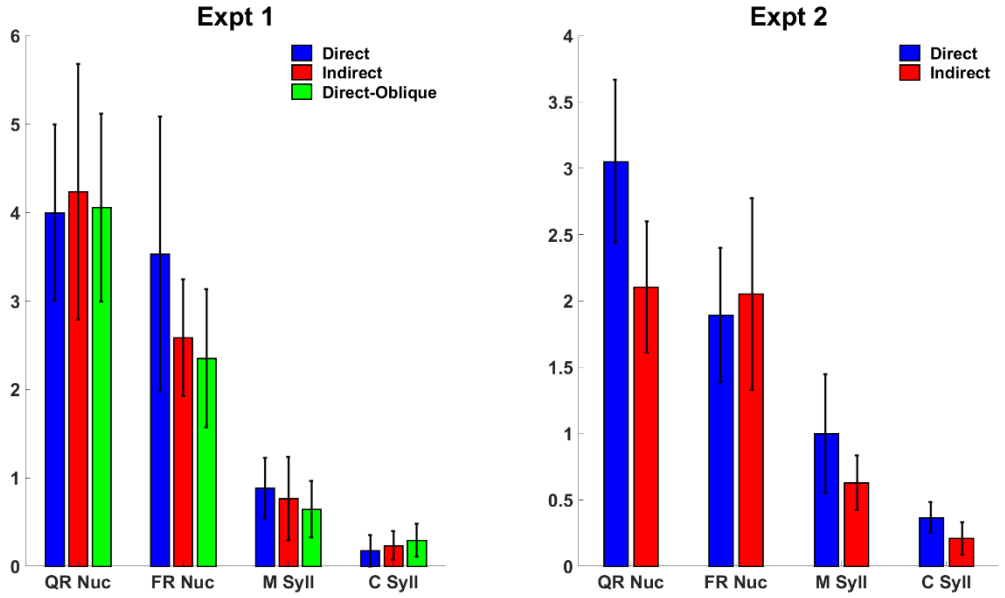

Figure S1. Mean number of vocalisations in each category for Expt 1 (left) and Expt 2 (right). QR Nuc = Quasi-Resonant nucleus, FR Nuc = Fully-Resonant nucleus, M Syll = Marginal syllable; C syll = Canonical syllable. Error bars indicate the standard error of the mean.

2.2 Correlations with neural coupling

Table S4 shows the correlation between adult-to-infant and infant-to-adult GPDC values (averaged across Theta and Alpha bands) and vocalisation duration, for each Experiment. As there was no significant infant-to-adult sending in Experiment 1, these correlations were not computed. Infants' vocalisation duration was only correlated to their own neural sending patterns (i.e. infant-to-adult), and not the adults' sending patterns. Thus, infants were not vocalising for longer in response to the adult, rather, their longer vocalisations were having a stronger synchronizing effect *on* the adult. Since the analysed EEG segments excluded periods of infant vocalisations (motion), speech artifacts could not account for this effect. Further, the neural-vocalisation relationship emerged only under Direct gaze from the adult, and was absent during Indirect gaze, consistent with the availability of the adult providing a stimulus for infants to vocalise with stronger communicative intent toward her. There were no significant correlations between neural connectivity and *number* of vocalisations for any gaze condition. This suggests that not every vocalization was equally effective in increasing neural connectivity with the adult. Rather, sustained vocalisations of a longer duration were more effective in influencing the adult.

	Gaze Condition	Adult-to-Infant **r** *(raw p-val)*	Infant-to-Adult **r** *(raw p-val)*
Expt 1	Direct	**-0.46** *(.11)*	*N.A.*
	Indirect	**0.03** *(.93)*	*N.A.*
	Direct-Oblique	**-0.25** *(.41)*	*N.A.*
Expt 2	Direct	**-0.09** *(.47)*	***0.67** *(.00)*
	Indirect	**-0.12** *(.28)*	**0.07** *(.78)*

*Table S4. Pearson correlation r-values and raw (uncorrected) p-values for adult-to-infant and infant-to adult connectivity (GPDC averaged over Theta and Alpha) and infant vocalisation duration in each gaze condition. *p<.05 (Benjamini-Hochberg FDR corrected)*

3 Nursery rhyme stimuli

3.1 Experiment 1 : Video

	Duration (s)			Mean Pitch (Hz)			Pitch Variability (Hz)			Loudness (dB)		
	Direct gaze	Indirect gaze	Direct-Oblique gaze	Direct gaze	Indirect gaze	Direct-Oblique gaze	Direct gaze	Indirect gaze	Direct-Oblique gaze	Direct gaze	Indirect gaze	Direct-Oblique gaze
If You're Happy	14.05	13.92	13.78	261.6	261.9	261.1	34.9	37.6	36.3	70.0	70.0	70.0
Hickory Dickory Dock	6.84	6.76	6.93	224.2	224.4	224.4	39.3	33.0	33.7	70.0	70.0	70.0
Humpty Dumpty	7.58	7.75	7.61	211.4	211.3	211.8	25.0	24.2	23.3	70.0	70.0	70.0
Old MacDonald	19.29	19.11	19.33	246.5	246.2	246.8	36.7	35.4	36.4	70.0	70.0	70.0
Where is Thumbkin	13.32	13.48	13.14	257.5	258.4	257.4	53.7	53.3	49.9	70.0	70.0	70.0
Twinkle Twinkle	21.04	20.87	21.11	245.5	245.9	245.6	37.2	35.5	37.5	70.0	70.0	70.0
Wheels on the Bus	10.54	10.6	10.8	243.6	243.2	243.3	42.0	43.8	41.2	70.0	70.0	70.0
Average (SD)	***13.24 (5.45)***	***13.21 (5.36)***	***13.24 (5.43)***	***241.47 (17.85)***	***241.61 (18.03)***	***241.49 (17.60)***	***38.41 (8.60)***	***37.55 (9.08)***	***36.90 (8.00)***	***70.0 (-)***	***70.0 (-)***	***70.0 (-)***

Table S5. Acoustic parameters of pre-recorded video nursery rhyme stimuli used in Experiment 1. Note that the loudness of all stimuli was equalized to 70dB. Pitch variability was computed as the standard deviation of pitch values across all timepoints in each stimulus.

Repeated Measures ANOVA for **Duration** : $F(2,12) = .10$, $p = .91$ (n.s.)
Repeated Measures ANOVA for **Mean Pitch** : $F(2,12) = .38$, $p = .69$ (n.s.)
Repeated Measures ANOVA for **Pitch Variability** : $F(2,12) = 1.31$, $p = .31$ (n.s.)

3.2 Experiment 2 : Live

	Mean Duration (s)			Mean Pitch (Hz)			Pitch Variability (Hz)			Loudness (dB)		
	Direct gaze	**Indirect gaze**	***t-test p-val*** *(BH-corr)*	**Direct gaze**	**Indirect gaze**	***t-test p-val*** *(BH-corr)*	**Direct gaze**	**Indirect gaze**	***t-test p-val*** *(BH-corr)*	**Direct gaze**	**Indirect gaze**	***t-test p-val*** *(BH-corr)*
If You're Happy	13.84 (0.54)	13.88 (0.49)	*0.77*	264.58 (16.64)	260.94 (21.28)	*0.77*	39.97 (8.43)	39.42 (9.24)	*0.77*	56.46 (1.92)	56.18 (1.74)	*0.77*
Hickory Dickory Dock	7.34 (0.43)	7.31 (0.51)	*0.84*	232.88 (19.33)	234.84 (21.95)	*0.77*	39.67 (11.34)	41.04 (10.82)	*0.77*	54.19 (2.97)	53.96 (3.16)	*0.77*
Humpty Dumpty	8.70 (0.61)	8.63 (0.63)	*0.77*	206.01 (8.54)	207.36 (5.99)	*0.77*	37.87 (8.07)	38.45 (6.12)	*0.77*	54.89 (3.41)	54.76 (3.66)	*0.77*
Old MacDonald	19.63 (1.08)	19.49 (1.29)	*0.77*	242.17 (14.36)	243.77 (18.96)	*0.77*	39.45 (7.32)	39.41 (7.57)	*0.98*	54.12 (3.18)	53.88 (3.04)	*0.77*
Where is Thumbkin	13.19 (0.73)	13.23 (0.62)	*0.77*	252.83 (17.97)	252.84 (16.67)	*0.99*	57.21 (9.91)	57.75 (9.66)	*0.77*	54.72 (3.32)	54.85 (3.31)	*0.77*
Twinkle Twinkle	21.62 (1.10)	21.39 (0.98)	*0.77*	248.56 (13.35)	246.46 (14.36)	*0.77*	42.26 (8.07)	43.04 (8.67)	*0.77*	52.72 (2.84)	52.54 (3.36)	*0.77*
Wheels on the Bus	11.31 (0.53)	11.28 (0.40)	*0.77*	261.44 (12.20)	262.09 (14.45)	*0.77*	47.20 (6.46)	46.58 (6.80)	*0.77*	54.85 (2.77)	54.66 (2.78)	*0.77*

Table S6. Acoustic parameters of live nursery rhyme stimuli used in Experiment 2.

Nursery rhymes were videoed live and the timings analysed post hoc. For each nursery rhyme, the average duration, mean pitch, pitch variability and loudness during Direct and Indirect conditions is given, and the SD is shown in brackets. Paired sample t-tests were calculated to assess whether the average duration, mean pitch, pitch variability or loudness of any of the nursery rhymes was significantly different across gaze conditions. No significant differences for any acoustic parameter or nursery rhyme were identified at the Benjamini-Hochberg (BH) FDR-corrected threshold of $p<.05$.

4 Experiment 2 : Experimenter's gaze perspective

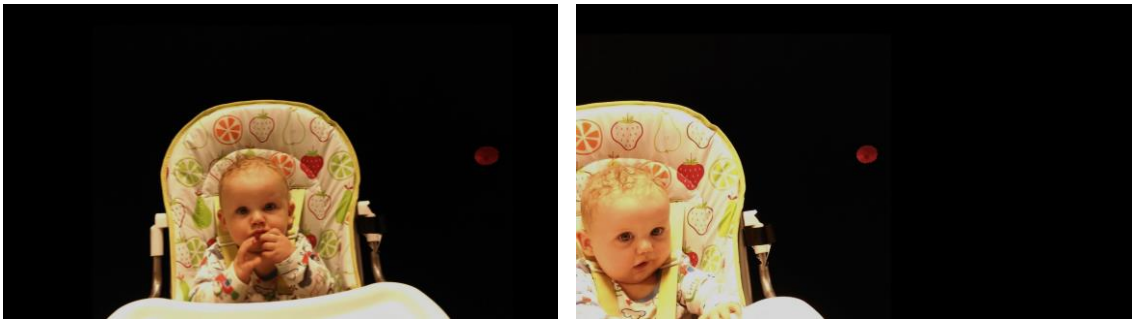

Figure S2. Example of experimenter's view during Direct gaze (left) and Indirect gaze (right)

During Direct gaze, the experimenter fixated on the infant and during Indirect gaze, she fixated on a red visual target placed 20º to the right or left side of the infant (in Figure S2, the target is placed on the right). Note that even during Indirect gaze, the infant was still clearly visible in her visual field. The infants' image was more peripheral and also very slightly larger during Indirect gaze because by rotating her head, she also brought her contralateral eye slightly closer to the infant.

Of note, in Experiment 2, we observed some, but much-reduced, infant-to-adult coupling in the Indirect condition. This was not unexpected, since the infant was facing the adult directly in both conditions, and, for the adult, the infant was positioned at 20° eccentricity from the fixation point, and so still clearly visible when her gaze was averted.

5 EEG acquisition

In both experiments, EEG signals were acquired using wireless amplifiers to reduce distraction for the infant during testing. In Experiment 1, EEG signals were obtained using a 32-channel wireless Biopac Mobita Acquisition System and 32-channel Easycap caps with electrodes placed at Fp1, Fp2, AFz, Fz, F3, F4, F7, F8, FC1, FC2, FC5, FC6, T7, T8, FT9, FT10, Cz, C3, C4, CP1, CP2, CP5, CP6, Pz, P3, P4, P7, P8, TP9, TP10, POz and Oz according to the International 10–20 placement system. EEG was recorded at 500 Hz with no online filtering using AcqKnowledge software (Biopac Systems Inc). In Experiment 2, EEG signals were recorded from C3 and C4 locations at 1000 Hz using a 2-channel Biopac MP150 Acquisition System with filters set at 0.1 Hz highpass and 100 Hz lowpass using AcqKnowledge software (Biopac Systems Inc). Both adult and infants' data was recorded concurrently in a single acquisition session on the same computer, ensuring accurate time synchronisation of the two data streams.

Prior to electrode or cap attachment, electrode sites were marked and wiped with alcohol. Conductive electrode gel was used to affix the electrodes/cap to the scalp. In Experiment 2, EEG was recorded from central sites to reduce potential confounding influences of muscle artefacts and blinking while still capturing a robust neural response (see analysis of speech production artifacts in Section 7). Across both experiments, a vertex reference location was used because it produces comparable results to other reference sites [4], and is the least invasive for young infants.

6 EEG artifact rejection

To ensure that the EEG data used for analysis reflected only attentive and movement-free behavior we performed a two-stage artifact rejection procedure. First, each experimenter-infant dyad was video-taped and the videos were reviewed frame-by-frame (30 fps) to identify the onset and offset times of movement artifacts, including blinks, head and limb motion, and chewing. Only periods when infants were still and looking directly at the experimenter were accepted. Next, manual artifact rejection was performed on this still, attentive data to further exclude segments where the amplitude of infants' or adults' EEG exceeded +100 μV.

6.1 Experiment 1 : Video

Following the two-stage artifact detection and rejection process, 17/19 infants (12M/5F), gave sufficient data for inclusion in the final analysis. The median (st. err.) age of the retained infants was 8.0 months (0.28 months). On average, the retained infants contributed 94.82 seconds (range = 25s to 171s, SD = 43.02s) of attentive and artifact-free data in the Direct gaze condition, 86.29 seconds (range = 31 to 169s, SD = 41.67s) in the Indirect gaze condition and 89.00 seconds (range = 35 to 170s, SD = 43.08s) in the Direct-Oblique gaze condition. Adult data was only analysed for those segments in which the infant data was retained. A larger quantity of clean and attentive data was obtained in the Direct gaze condition than in the Indirect gaze condition ($t(16) = 2.83$, $p < .05$), but data quantity did not differ between Direct-Oblique and Indirect gaze conditions ($t(16) = 1.16$, $p=.26$) or between Direct and Direct-Oblique gaze conditions ($t(16) = 1.54$, $p=.14$). However, additional analyses performed to assess the effect of these data quantity differences (e.g. sub-sampling an equal number of epochs across gaze conditions) confirmed that the main effects of gaze were not affected by data quantity.

6.2 Experiment 2 : Live

Following artifact rejection, 19/29 infants (10M/9F), gave sufficient data for inclusion in the final analyses. The median (st.err.) age of retained infants was 8.52 (0.57) months. On average, the retained infants contributed 45.52 seconds (range = 8s to 107s, SD = 28.18s) of attentive and artifact-free data in the Direct gaze condition, and 43.92 seconds (range = 11 to 123s, SD = 30.07s) in the Indirect gaze condition. A paired t-test confirmed that there was no significant difference in the amount of clean data obtained between Direct and Indirect gaze conditions ($t(18) = 0.44$, $p = .66$) therefore all the clean data was used for analysis. Adult data was only analysed for those segments in which the infant data were retained.

7 Adult speech artifact analysis (speaking versus rest)

Speech production artifacts were present in the EEG signal of the adult speaker, and these articulatory motions are known to reduce the signal-to-noise ratio of neural signals that relate to cognition [5]. For instance, the temporalis muscle is used for closing the lower jaw and this muscle spreads widely over the scalp locations that correspond to the frontal/temporal/parietal junction of the brain, generating large artifacts in the EEG signals measured over these regions [5]. Muscle artifact contamination is greatest over frontal and temporal scalp regions [6] and generally less severe over central regions, where our recording electrodes were placed. Several methods have been proposed for removing speech artifacts

from the EEG signal. These include the use of low-pass filtering to remove muscle artifacts that most prominently occur at frequencies over 12 or 20 Hz [7,8], and blind source separation based on Canonical Correlation Analysis [6] or Independent Component Analysis [9] to separate cortical sources from electromyographic (EMG) responses. However, none of these methods are able to completely remove motion artifacts from the EEG signal, and may even remove some genuine neural activity of interest.

Therefore, in order to understand whether these speech production artifacts could have introduced a pattern of bias into our results, it is first necessary to quantify the spatial (i.e. scalp topography) and spectral signature of the exact speech production artifacts that were generated by the adult speaker whilst singing nursery rhymes. According, we performed a control analysis to systematically document the topographical and spectral differences in the EEG signals of the speaker during speech production (in each gaze condition) as compared to rest.

7.1 Protocol

All recordings were performed by the same female speaker as in the main studies.

Nursery rhymes. Twenty repetitions of each of the 7 nursery rhymes were recorded by the speaker in each of three gaze positions (Direct, Indirect and Direct-Oblique), in which the speaker maintained the same head and body position as in the original experiments. During recording, her gaze was fixated on a life-sized head image of an infant.

Resting State. The adult was instructed to remain relaxed with her eyes open and to focus her gaze on the image of the infant. She was told to avoid eye, head or other movements. Resting state EEG was recorded for 12 minutes.

7.2 EEG acquisition

32-channels of EEG data were acquired from the adult at 500 Hz using a Biopac Mobita amplifier and Acqknowledge v5.0 software. No online referencing or filtering was used. Impedance for all channels was under 10KΩ.

7.3 EEG pre-processing and analysis

Average re-referencing was performed offline. No filtering was applied to the raw signal. Eye-movement and blink artifacts, as well as segments with raw amplitude above 100 μV were manually identified and removed from the raw recordings. After cleaning, the 20 repetitions of each nursey rhyme were concatenated for each gaze condition. A Fast Fourier Transform was applied to the nursery rhyme and resting state data in non-overlapping 1.0s windows for each EEG channel. As the frequency spectra of individual nursery rhymes did not differ, we collapsed the data across nursery rhymes and analysed the grand average frequency spectrum over all nursery rhymes.

7.4 Scalp topography during resting state and speech production

The scalp topography of EEG power in 5 frequency bands (Delta[1-3Hz]; Theta[3-6Hz]; Alpha[6-9Hz]; Beta[9-25Hz]; Gamma[25-42Hz]) is shown in Figure S3 for resting state condition, and during speech production for each gaze condition.

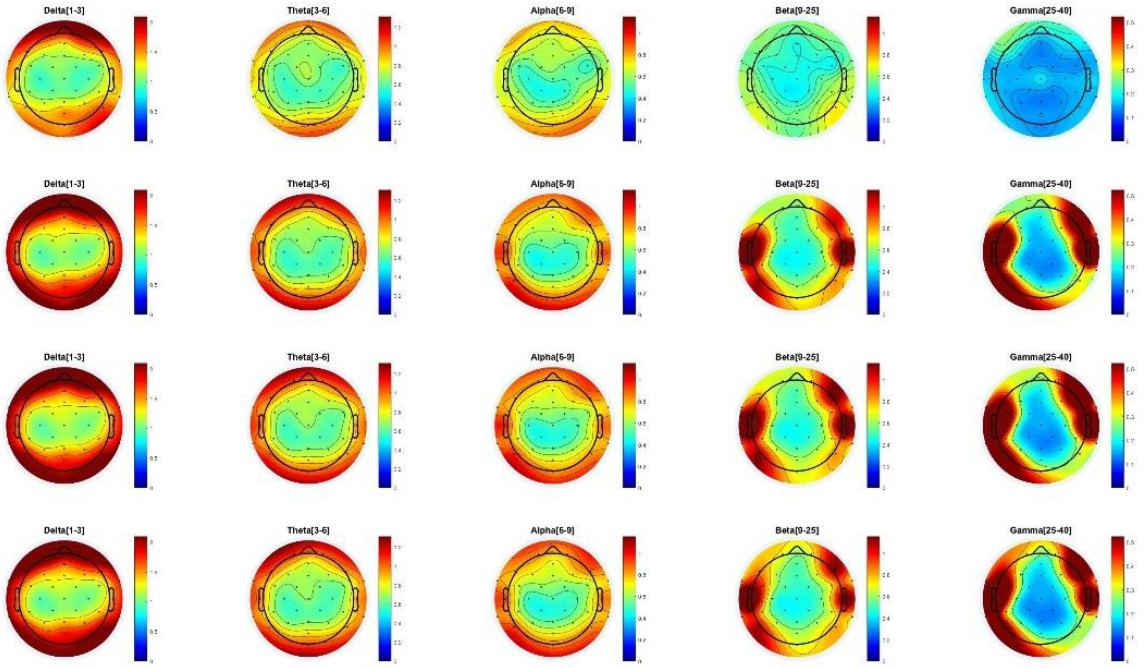

Figure S3. Scalp topography of EEG power in Delta, Theta, Alpha, Beta and Gamma bands for Resting State (top row), Direct Gaze (second row), Indirect Gaze (third row) and Direct-Oblique Gaze (bottom row). Color scaling for each topographical plot is identical.

From visual inspection, it may be observed that during speech production (as compared to resting state), there were distinct increases in power, especially at Beta and Gamma frequencies, and particularly over left and right fronto-temporal regions. However, central regions (e.g. C3 and C4) appeared to be the least affected by speech production power artifacts. To assess these differences more closely, a detailed spectral analysis on the power spectrum at C3 and C4 was performed to test for frequency-specific changes in power during speech production as compared to resting state, as described next.

7.5 Spectral analysis at C3 and C4

To identify spectral differences between speech production conditions and resting state, one-way ANOVAs with 4 levels (RS, Direct gaze, Indirect gaze, Direct-Oblique Gaze) were conducted at each frequency between 0 to 40 Hz, for C3 and C4 channels. We performed all post-hoc comparisons (RS vs each gaze condition; each gaze condition against the other two gaze conditions) by running unpaired t-tests. For all tests, boot-strapping was performed by randomly selecting an equal subset (~1500) of 1.0s segments in each gaze condition, and permutating this selection 100 times. Only comparisons in which t-tests were significant at the alpha-level of $p<0.05$ for over 95% of all permutations are reported.

As shown in Figure S4, the results of the ANOVA revealed that there were significant spectral differences between speaking and rest conditions at C3 at 12 Hz, 13 Hz and between 15 – 39 Hz. At C4, significant differences were observed between 21 – 25 Hz, and between 29 - 39 Hz. In each case, speech production *increased* power in the EEG signal relative to

rest. Note that for both electrodes, *no overall* differences in power were observed across conditions at frequencies under 12 Hz.

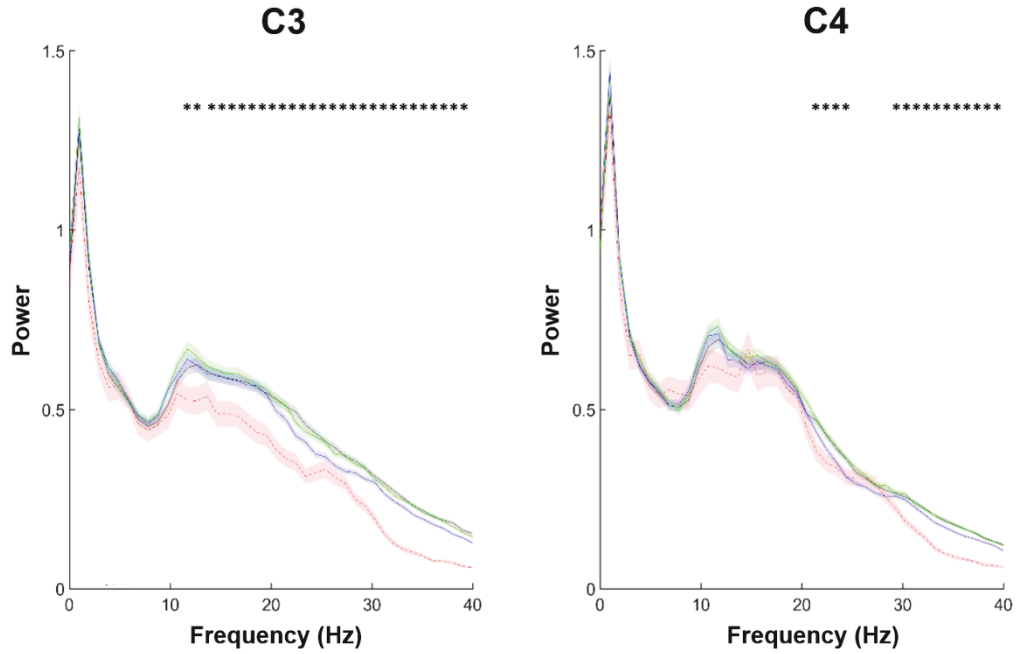

Figure S4. Power spectrum and results of ANOVA conducted at C3 (left plot) and C4 (right plot) for spectral differences between the 4 conditions (RS [red line], Direct Gaze [black line], Indirect Gaze [green line], and Direct-Oblique Gaze [blue line]). The x-axis indicates Frequency (Hz) and the y-axis indicates power (units). Shaded areas indicate the 95% confidence interval around the mean for each condition, and [*] indicates an overall difference between conditions in the ANOVA analysis at p<.05.

To assess the specific pattern of differences between conditions, post-hoc t-tests were conducted at every frequency between 0 and 40 Hz (as described above). For C3, significant differences between Rest and Direct gaze were observed at 12,13 and 15 – 39 Hz. For Indirect gaze, differences were additionally observed in the Delta band at 2 Hz. For Direct-Oblique gaze, differences were observed at 12, 13, 15 – 23 and 28 – 39 Hz. For C4, Direct gaze differed from Rest at 21 – 23 and 29 - 39 Hz. For Indirect gaze, differences were observed at 21 – 25 and 29 – 39 Hz. For Direct-Oblique gaze, differences were only observed between 29 – 39 Hz. **Over all comparisons, no significant differences between speaking and rest were observed in Theta (3-6 Hz) and Alpha (6-9 Hz) bands.**

In summary, our analysis of the adults' speech production artifacts confirmed that speech gestures did indeed produce increases in EEG power that were most prominent over frontal and temporal scalp regions, consistent with previous studies [5,6]. However, our fine-grained spectral analyses revealed that, relative to resting state EEG (when no overt motor activity was present), electrodes in the central scalp region (C3 and C4) showed no significant change in power at Theta (3 – 6 Hz) and Alpha (6 – 9 Hz) band frequencies for any gaze condition. Accordingly, our main connectivity analyses focused on this scalp region and frequency range.

8 EEG power analysis

As our main aim was to assess changes in connectivity between gaze conditions, it was important to first establish whether there were any properties of the underlying EEG

signal in each condition that might artifactually generate increases (or decreases) in computed connectivity. One such potential confounding factor is the composition of the power spectrum of the EEG signal. The accuracy of the partial directed coherence (PDC) metric can be sensitive to even moderate changes in signal-to-noise ratio [10]. For example, Adhikari et al [10] reported that a 10% decrease in signal power from 67% to 57% was associated with ~15% lower accuracy in PDC directionality estimation, although a similar 11% power change from 57% to 46% only caused an accuracy drop of <5%. Therefore, if the EEG signal in one experimental condition has higher noise than in another condition (or if the spectral composition of the signal changes substantially), this can lead to greater error in estimation of connectivity patterns.

To assess the power spectra of the EEG signals, their power spectral density (PSD) was estimated using the Matlab *'periodogram.m'* function, which performs a discrete Fourier transform on the signal. One PSD estimate was computed for each channel (left and right electrodes for adult and infant respectively), for each participant pair, and for each experimental condition. The resulting power spectra were then divided into EEG frequency bands, and averages were taken for each frequency band and used for analysis.

To assess whether there were differences in EEG power between the gaze conditions, for each experiment, a repeated measures ANOVA was conducted taking Gaze ([3 (Expt 1) or 2 (Expt 2) levels]), Frequency band ([2 levels, Theta 3-6 Hz and Alpha 6-9 Hz]) and Channel ([4 levels, infant and adult x left and right) as within-subjects factors.
For Experiment 1, there was no overall difference in EEG power between the Direct, Indirect and Direct-Oblique conditions ($F(2,32) = 0.25$, $p = .78$). There was also no interaction between Gaze x Channel ($F(6,96) = .23$, $p = .97$), no interaction between Gaze x Frequency ($F(2,32) = 1.94$, $p = .16$), and no interaction between Gaze x Channel x Frequency ($F(6,96) = 2.04$, $p = .07$). For Experiment 2, there was again no overall difference in EEG power between the Direct and Indirect conditions ($F(1,18) = 0.30$, $p = .59$). There was no interaction between Gaze x Channel ($F(3,54) = .14$, $p = .93$), no interaction between Gaze x Frequency ($F(1,18) = .00$, $p = .98$), and no interaction between Gaze x Channel x Frequency ($F(3,54) = .90$, $p = .45$). Therefore, the gaze manipulation did not generate any detectable power changes that might systematically bias the PDC metric.

9 Neural connectivity analysis : Partial Directed Coherence (PDC)

Partial Directed Coherence (PDC) is a directional causal measure of direct flows between channels [11-13]. It is based on the principles of Granger Causality [14], and measures the degree of influence that channel j (the 'Sender') *directly* has on channel i (the 'Receiver') with respect to the total influence of j on all channels in the network. Here, each individual electrode (Infant L, Infant R, Adult L, Adult R) was taken as one channel and the entire network consisted of 4 electrodes in total. We computed directed coherence values for all 12 possible pairwise connections, both within individual (e.g. Infant L -> Infant R) as well as across individuals (e.g. Infant L -> Adult L).

For the current analysis, we used Generalised Partial Directed Coherence (GPDC; [12]), which is an adapted version of PDC with better variance stabilization properties and the advantage of scale-invariance [15]. As a first step in the analysis, a multivariate autoregressive (MVAR) model is fitted to the EEG time series, which has the advantage of providing information about causal linear interaction effects in addition to estimating the coupling strength between channels. A frequency representation of the MVAR model parameters is then generated via a Fourier Transform, as follows:

$$A(f) = I - \sum_{p=1}^{P} A_p e^{-2\pi i p(\frac{f}{f_s})} \quad \text{(eq.1)}$$

where A_p are the model coefficients, I refers to the M-dimensional identity matrix, *fs* is the sampling frequency, and $i^2 = -1$. For each pair of channels (*i* and *j*), $GPDC_{ij}$ is then computed as :

$$GPDC_{ij}(f) = \frac{\frac{1}{\sigma_i} |A_{ij}(f)|}{\sqrt{\sum_{m=1}^{M} \frac{1}{\sigma_m{}^2} |A_{mj}(f)|^2}} \quad \text{(eq.2)}$$

where σ_i^2 refers to the variance of the innovation process $x_i(t)$. GPDC takes values between [0,1] and is normalized across receivers (i.e. total outflow = 1 at each frequency), with larger values indicating strong connectivity.

The MVAR model was estimated using the Burg-type Nuttall-Strand method [16] which is thought to perform best for small sample sizes [17], and a model order (MO) of 5 was used. The model order (MO) indicates the number of preceding samples that are used to predict the data at sample time *t*, and determines the number of observed frequency components for each pair of channels, which is typically half the model order. Following prior studies on autoregressive modeling [18,19] and multivariate autoregressive modeling of EEG time series [20-23], here a model order of 5 was used for this analysis. For example, Jansen et al [18] reported that a fifth order AR model was sufficient in 90% of cases to adequately capture variance in EEG time series data. Vaz et al [19] also noted that *"a 5th order AR model represents adequately 1- or 2-s EEG segments with the exception of featureless background, where higher order models are necessary"*. Model orders used in other MVAR EEG studies typically range between 3 and 6 [20-23].

One MVAR model and the resulting set of GPDC estimates (spanning the entire frequency spectrum) was computed for each non-overlapping 1.0s EEG epoch (200 data samples), and these estimate GPDC values were averaged across all epochs for each participant pair, for each experimental condition. The resulting epoch-averaged GPDC spectrum was then divided into discrete Theta (3-6 Hz) and Alpha (6-9 Hz) EEG frequency bands. Note that as infants' Theta and Alpha EEG bands are lower in frequency as compared to adults [24], our frequency banding was adjusted lower accordingly. The mean GPDC value was taken within each frequency range, for each pairwise connection, condition and participant.

10 Control analysis 1: Surrogate connectivity data

As a control analysis, we generated a surrogate dataset comprising 1000 temporally-shuffled versions for each participant pair. The aim of this control analysis was to disrupt the fine-grained temporal correspondence between adult and infant neural signals by randomly pairing each adult 1.0s epoch with a non-matching infant 1.0s epoch from a different timepoint within the same experimental session. For example, the adult neural signal whilst singing "Twinkle Twinkle" may be paired to the infant signal whilst listening to "Wheels on the Bus". The pairing of adult and infant time-shuffled epochs was determined by random permutation, and was non-identical for each of the 1000 shuffled versions generated for each participant pair, as well as for different participant pairs.

This shuffled control allowed us to establish a baseline level of non-specific connectivity between brains that could have arisen, for example, from commonalities in the physical environment during a particular experimental session, or due to general increases in infants' and adults' arousal. We could then be assured that any neural connectivity which could be detected over and above this baseline was specifically related to the time-contingent neural coupling between speaker and listener for the given experimental stimulus. Identical connectivity analyses were then performed on the real and surrogate datasets. All GPDC analyses were performed using the eMVAR (Extended Multivariate Autoregressive Modelling) Toolbox [25] in Matlab (The Mathworks Inc). The resulting GPDC values are shown in Tables S1 and S2.

11 Control analysis 2: Neural entrainment to the speech stimulus

In order to assess whether interpersonal connectivity gaze effects could be attributed to differences in basic speech processing across gaze conditions, we examined whether neural oscillatory entrainment to the amplitude envelope (temporal structure) of the adult's speech signal differed between gaze conditions. The EEG data was first low-pass filtered under 45 Hz using an inverse fft filter to remove line noise (EEGLAB *eegfiltfft.m* function [26]). Next, the wholeband amplitude envelopes of the speech signal of the nursery rhyme stimuli were extracted using the Hilbert transform. To assess the degree of entrainment between the neural EEG signal and the speech amplitude envelope the phase-locking value (PLV, [27]) was computed. The PLV takes values between [0, 1], where a value of 0 reflects the absence of phase synchrony and a value of 1 reflects perfect synchronisation.

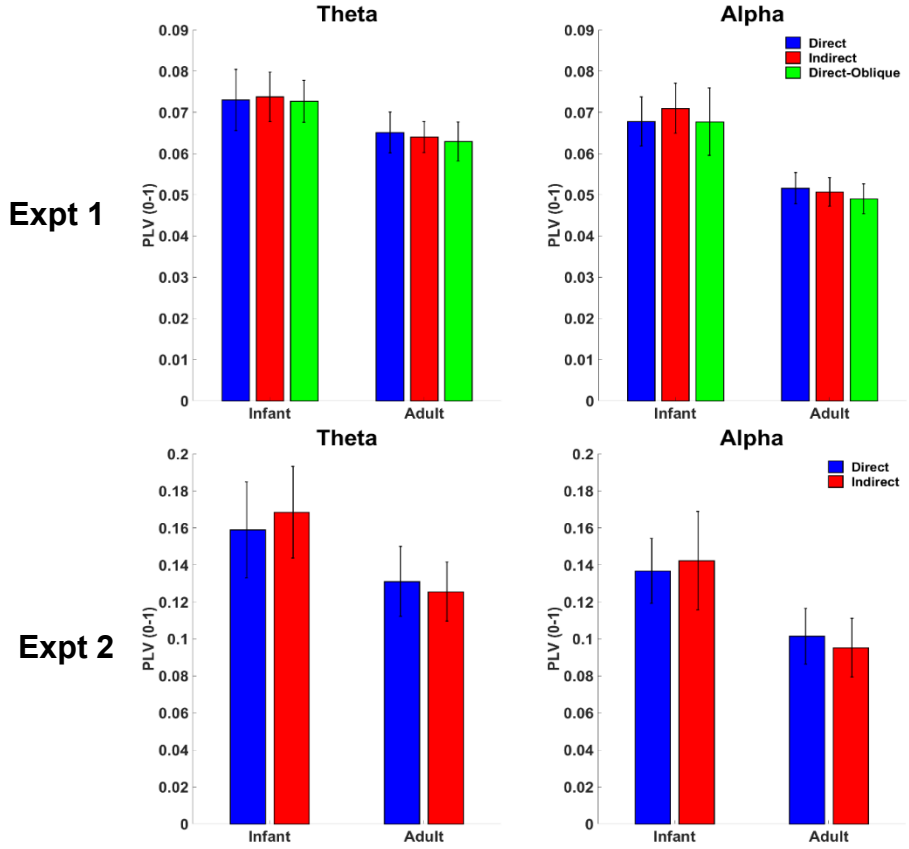

Figure S5. Speech-brain entrainment for Experiment 1 (top panel) and Experiment 2 (bottom panel) infants and adults by gaze condition for Theta (left) and Alpha (Right) frequency bands respectively. Error bars show the standard deviation.

Prior to calculating the PLV, a continuous wavelet transform was applied to the neural and speech data, which convolves each time series with scaled and translated versions of a wavelet function [28]. Here, the wavelet function chosen was the complex Morlet wavelet (bandwidth of mother wavelet = 1 Hz, time resolution = 0.1 Hz). The wavelet time-frequency decomposition was performed at 40 log-spaced frequencies. The phase series at each frequency was extracted from the complex wavelet coefficients, and divided into matching EEG and speech epochs of length 2.0s (with no overlap). The PLV for each epoch was then computed, and averaged over all epochs for each participant. Finally, frequency band-averaged PLV values were computed for Theta (3-6 Hz) and Alpha (6-9 Hz) frequency bands for each gaze condition, as shown in Figure S5.

For each experiment, a Repeated Measures ANOVA was conducted taking Gaze (3 or 2 levels) and Frequency (2 levels) as within-subjects factors, and Group (Infant or Adult) as the between-subjects factor. For both experiments, there was *no* significant difference in speech-brain entrainment between gaze conditions (Expt 1 : $F(2,64) = 1.89$, $p = .16$; Expt 2 : $F(1, 36)=.06$, $p=.80$), and no significant interaction between Gaze and Frequency (Expt 1 : $F(2, 64)=.72$, $p=.49$); Expt 2 : $F(1, 36)=.42$, $p=.52$), suggesting that gaze did not change the pattern of speech-brain entrainment for Theta or Alpha bands. Therefore, any interpersonal connectivity gaze effects cannot be attributed to differences in basic speech processing.

12 Full infant scalp topography of receiving GPDC values (Expt 1)

To assess the scalp topography of infants' neural receiving patterns with respect to adults' C3 and C4 electrodes, 4-channel GPDC analyses were conducted for all hemispherically-dichotomous pairs of infants' electrodes (e.g. infants' left temporal [T7] and right temporal [T8] electrodes). The results indicated that across both EEG frequency bands, and across all gaze conditions, the strongest adult-to-infant connectivity was observed over infants' central and posterior scalp locations (including C3 and C4). By contrast, lower connectivity was observed over infants' frontal and temporal regions, particularly for the Alpha band. This topographical pattern confirms that the connectivity data from C3 and C4 (reported in the main manuscript) is indeed representative of infants' overall neural response to the adult.

13 ANOVA results of gaze effects on interpersonal neural connectivity

To recap our analysis approach, the average (a) infant-to-adult GPDC *(I→A)* and (b) adult-to-infant *(A→I)* GPDC was computed for each gaze condition, and for Theta and Alpha bands separately. We then conducted Repeated Measures (RM) ANOVAs using these average indices, taking Frequency (2 levels) and Gaze (3/2 levels) as within-subjects factors. From Table S2, it may be noted that a few individual connections in Expt 2 were not significantly above threshold for one of the gaze conditions (but this never occurred across both gaze conditions). These values were included in the grand averages in order to maintain representativeness and a balanced ANOVA structure, since they were still statistically meaningful in terms of potentially revealing a difference *between* gaze conditions. For all analyses, infants' looking times across each gaze condition were entered as co-variates to control for individual differences in attentiveness. For the I→A analysis, age was entered as an additional co-variate to control for individual differences in infants' maturation. To assess the specific gaze effects at each frequency, we conducted planned pairwise comparisons using Dunnett's multiple range t-test [29], which independently controls for familywise error rate *without* a prior F-test. Requiring a significant F-test before performing multiple

comparison tests (like Dunnett's) is not recommended as this inflates the false negative rate [30,31]. At each frequency, we performed 3 planned pairwise comparisons using Dunnett's test : (1) Direct > Indirect (Expt 1 & Expt 2); (2) Direct-Oblique > Indirect (Expt 1 only) and (3) Direct = Direct-Oblique (Expt 1 only). The results of these pairwise tests are reported in the main text. Here we provide a breakdown of the RM ANOVA results.

13.1 Experiment 1 : Adult-to-infant GPDC (A→I)

RM ANOVA Effect	
Gaze	F(2,26) = .66, p = .53, $\eta^2p = .05$
Frequency	F(1,13) = 6.92, p<.05, $\eta^2p = .35$
Gaze x Frequency	F(2,26) = 1.61, p = .22, $\eta^2p = .11$

Table S7 – Experiment 1 adult-to-infant RM ANOVA results

It may be noted that although there were strong pairwise differences between individual gaze conditions (as revealed by Dunnett's test and reported in the main text), the overall F-test for the Gaze effect was not significant. This apparent discrepancy could arise from the fact that the null hypothesis for the ANOVA F-test is that the means across all gaze conditions (and frequencies) are equal. However, this null hypothesis is inconsistent with our a-priori predictions that Direct/Direct-Oblique gaze would both differ from Indirect gaze, but Direct gaze would *not* differ from Direct-Oblique gaze. Therefore, we expected 2 out of our 3 condition means to be equal, and only 1 to differ. Conversely, if we had conducted the ANOVA analysis with only 2 gaze conditions that were predicted to differ (such as Direct-Oblique versus Indirect), then there would indeed be a significant main effect of Gaze (F(1,14) = 6.05, p<.05, η^2p = .30). However, conducting 3 separate ANOVAs for each pairwise gaze contrast would be unparsimonius and lead to Type I error inflation. Accordingly, the ANOVA F-test was ill-suited to evaluate our predicted hypotheses in Experiment 1. To address this, we relied on the findings of the Dunnett's tests (which independently control for Type 1 error) to assess our specific predictions in Experiment 1.

13.2 Experiment 2 : Adult-to-infant GPDC (A→I)

RM ANOVA Effect	
Gaze	F(1,16) = 5.51, p<.05, $\eta^2p = .26$
Frequency	F(1,16) = .00, p=.96, $\eta^2p = .00$
Gaze x Frequency	F(1,16) = 5.48, p<.05, $\eta^2p = .26$

Table S8 – Experiment 2 adult-to-infant RM ANOVA results

As expected, the results of the RM ANOVA showed a significant main effect of Gaze (Direct > Indirect), which corroborated with findings from pairwise Dunnett's tests (reported in the main manuscript) showing that *A→I* connectivity was higher for Direct > Indirect gaze in both Theta and Alpha bands.

13.3 Experiment 2 : Infant-to-adult GPDC (I→A)

RM ANOVA Effect	
Gaze	F(1,15) = 6.18, p<.05, $\eta^2p = .29$
Frequency	F(1,15) = 38.8, p<.001, $\eta^2p = .72$
Gaze x Frequency	F(1,15) = 10.5, p<.01, $\eta^2p = .41$

Table S9 – Experiment 2 infant-to-adult RM ANOVA results

For infant-to-adult connectivity in Experiment 2, the results of the RM ANOVA showed a significant main effect of Gaze (Direct > Indirect), which corroborated with findings from pairwise Dunnett's tests (reported in the main manuscript) showing that *I→A* connectivity was higher for Direct > Indirect gaze in both Theta and Alpha bands.

14 Infant looking times

14.1 Experiment 1 : Video

For Direct gaze stimuli, infants' average looking time was 101.61s (SD = 43.04s). Their looking time was 92.73s (SD = 41.73s) for Indirect gaze stimuli, and 95.13s (SD = 43.02s) for Direct-Oblique gaze stimuli. A repeated measures ANOVA analysis with Gaze (3 levels) as the within-subjects factor revealed that there was a significant main effect of Gaze ($F(2,32) = 4.46$, $p<.05$) on infants' looking times. Tukey HSD post hoc analysis indicated that infants looked for significantly longer at the Direct gaze nursery rhymes as compared to the Indirect gaze stimuli ($p<.05$), but there was no difference in looking time between Direct gaze and Direct-Oblique gaze ($p=.10$) or between Indirect gaze and Direct-Oblique gaze ($p = .72$). As the acoustic parameters of the video stimuli were tightly controlled across conditions, these differences in infants' looking patterns could not have arisen from inconsistencies in the speakers' presentation of the stimuli.

14.2 Experiment 2 : Live

For Direct gaze stimuli presented in a live format, infants' mean looking time was 61.01s (SD = 31.61s) and for Indirect live stimuli, infants' mean looking time was 61.11s (SD = 34.21s). A paired t-test confirmed that there was no significant difference in infants' looking time for Direct and Indirect gaze conditions ($t(18) = 0.03$, $p = .98$). Therefore, infants were not more inattentive during Indirect gaze for live stimuli.

It is interesting that infants showed a different pattern of looking for Direct versus Indirect gaze stimuli across the two experiments. In Experiment 1 (video), consistent with previous screen-based studies (Farroni et al, 2002), infants looked longer at Direct gaze than Indirect gaze stimuli. However, in Experiment 2 (live), infants looked equally long at both types of gaze stimuli. This apparent attentional benefit for live speech was also observed in a phonetic learning experiment by Kuhl et al (2003), in which infants were more attentive to (and showed more phonetic learning from) live adult speakers than DVD movies of the same speakers. However, even though infants were equally attentionally-engaged for Direct and Indirect gaze stimuli in Experiment 2, their neural connectivity to the adult differed across gaze conditions, suggesting that attention did not underlie the neural gaze effect.

15 Effect of infant age

We examined the effect of age based on a median split analysis that divided our data into younger and older infants (Experiment 1 = 8.0 months, Experiment 2 = 8.52 months), entering this as an additional between-subjects factor in the RM ANOVA analyses. For both Expt 1 and 2, there was no main effect of Age on adult-to-infant GPDC (Expt 1 : $F(1,12) = 1.38$, $p=.26$, $\eta^2p = .10$; Expt 2 : $F(1,15) = .00$, $p=.96$, $\eta^2p = .00$). There was also no significant interaction between Age and other factors (Frequency Band, Gaze, $p>.13$ for all). For infant-to-adult GPDC in Expt 2 (Alpha band), there was similarly no effect of Age ($F(1,15) = .15$, $p=.70$, $\eta^2p = .01$). Thus, the effects of Gaze did not differ as a function of infants' age.

16 Internal replicability of gaze findings

We conducted a permutation analysis to assess the internal replicability of our two main gaze findings ***[1] Direct > Indirect*** (E1 and E2) and ***[2] Direct-Oblique > Indirect*** (E1). In the permutation analysis, 71%/75% of the E1/E2 cohort data (N=12 or 14 out of 17 or 19) was randomly selected in all possible ways (=6,188 or 11,628 permutations). For each cohort permutation, one main test statistic was computed for each Gaze contrast. To permit direct comparison across Experiments 1 and 2, we selected the same test statistic for the Direct v Indirect contrast in each experiment : Alpha band adult-to-infant GPDC. For completeness and to avoid bias, a different frequency band was selected for the Direct-Oblique v Indirect contrast : Theta band adult-to-infant GPDC. For each permutation, an RM ANOVA was performed on the test statistic, taking Gaze as the within-subjects factor and controlling for infant looking time. The effect size (r^2) was recorded for each permutation to yield a distribution of possible effect sizes over all permutations.

15.1 Experiment 1

For the Direct vs Indirect gaze contrast (left subplot in Figure S6), the effect size (r^2) obtained across all permutations was 0.219 (mean) / 0.212 (median), indicating the presence of a medium-large effect size in the data. For the Direct-Oblique vs Indirect gaze contrast (right subplot, Figure S6), the effect size (r^2) obtained across all permutations was 0.192 (mean) / 0.183 (median), indicating the presence of a medium effect size in the data.

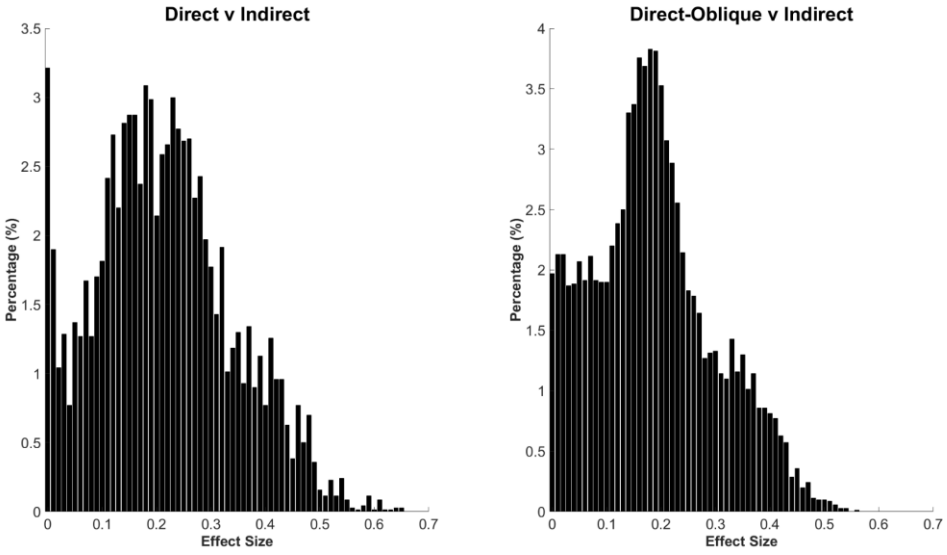

Figure S6. Experiment 1 : Distribution of effects sizes for Direct vs Indirect (left) and Direct-Oblique vs Indirect (right) contrasts obtained for 71% (n=12) sub-samples of the data.

15.2 Experiment 2

For the Direct vs Indirect gaze contrast (which used the same test statistic as Experiment 1), the mean effect size (r^2) obtained across all permutations was 0.332 (mean) / 0.321 (median), indicating the presence of a large effect size in the data (see Figure S7). It is also interesting to note that the effect size distribution appeared to be bimodal, which could indicate that there is a subset of infants who show particularly strong sensitivity to adult gaze.

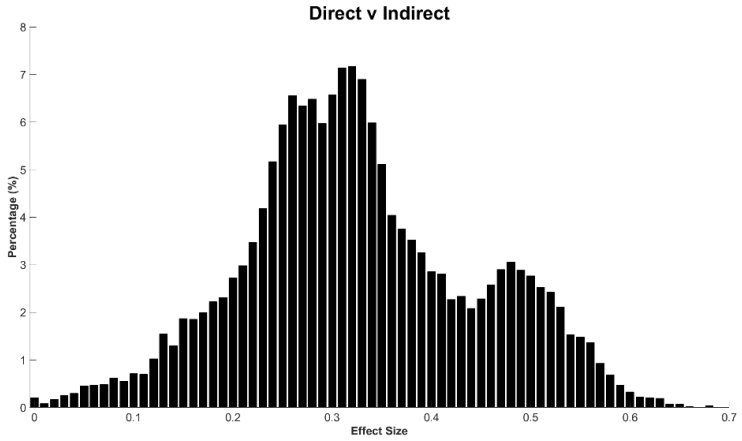

Figure S7. Experiment 2 : Distribution of effect sizes for the Direct vs Indirect contrast obtained for 75% (n=14) sub-samples of the data.

Comparing across Experiments 1 and 2, the Direct vs Indirect gaze contrast consistently yielded at least a medium-sized effect across both experiments, indicating that this gaze finding is replicable across two different testing modalities (video versus live presentation) and two different infant cohorts.

SI REFERENCES

[1] Oller, D.K. (2000). The emergence of speech capacity. Mahwah, NJ: Erlbaum.

[2] Kuhl, P. K., Tsao, F.-M., & Liu, H.-M. (2003). Foreign-language experience in infancy: Effects of short-term exposure and social interaction on phonetic learning. PNAS, 100(15), 9096–9101.

[3] Goldstein, M. H., & Schwade, J. A. (2008). Social feedback to infants' babbling facilitates rapid phonological learning. Psychological Science, 19(5), 515-523.

[4] Tomarken, A. J., Davidson, R. J., Wheeler, R. E., & Kinney, L. (1992). Psychometric properties of resting anterior EEG asymmetry: Temporal stability and internal consistency. Psychophysiology, 29(5), 576-592.

[5] Brooker, B.H., & Donald, M.W. (1980). Contribution of the speech musculature to apparent human EEG asymmetries prior to vocalisation. Brain and Language, 9, 226-245.

[6] De Vos, M., Riès, S., Vanderperren, K., Vanrumste, B., Alario, F.X, Van Huffel, S., Burle, B. (2010). Removal of muscle artifacts from EEG recordings of spoken language production. Neuroinformatics, 8, 135-150.

[7] Ganushchak, L.Y., & Schiller, N.O. (2008). Motivation and semantic context affect brain error-monitoring activity: an event-related brain potentials study. NeuroImage, 39, 395-405.

[8] Laganaro, M., & Perret, C. (2011). Comparing electrophysiological correlates of word production in immediate and delayed naming through the analysis of word age of acquisition effects. Brain Topogr., 24, 19-29.

[9] Porcaro, C., Medaglia, M.T., & Krott, A. (2015). Removing speech artifacts from electroencephalographic recordings during overt picture naming. NeuroImage, 105, 171-180.

[10] Adhikari, A., Sigurdsson, T., Topiwala, M.A., & Gordon, J.A. (2010). Cross-correlation of instantaneous amplitudes of field potential oscillations: a straightforward method to estimate the directionality and lag between brain areas. J Neurosci Methods, 191:191–200.

[11] Baccalá, L. & Sameshima, (2001). Partial directed coherence: a new concept in neural structure determination. Biological Cybernetics, 84(6), 463–474.

[12] Baccalá, L. A., Sameshima, K., & Takahashi, D. Y. (2007, July). Generalized partial directed coherence. In Digital Signal Processing, 2007 15th International Conference on (pp. 163-166). IEEE.

[13] Faes, L, & Nollo, G. (2010). Extended causal modelling to assess Partial Directed Coherence in multiple time series with significant instantaneous interactions. Biological Cybernetics, 103(5):387-400.

[14] Granger, C.W.J. (1969). Investigating causal relations by econometric models and cross-spectral methods. Econometrica, 37, 424-438.

[15] Hoerzer, G.M., Liebe, S., Schloegl, A., Logothetis, N.K., & Rainer, G. (2010). Directed coupling in local field potentials of macaque v4 during visual short-term memory revealed by multivariate autoregressive models. Frontiers in Computational Neuroscience, 4:14. doi: 10.3389/fncom.2010.00014

[16] Marple, S.L., & Nuttall, A.H. (1983). Experimental comparison of three multichannel linear prediction spectral estimators, IEE Proc. F 130, 218–229

[17] Schlogl, A. (2006). A comparison of multivariate autoregressive estimators. Signal Processing, 86:2426-2429.

[18] Jansen B.H., Bourne J.R., & Ward J.W. (1981) Autoregressive estimation of short segment spectra for computerized EEG analysis. IEEE Trans. Biomedical Engineering. 28(9).

[19] Vaz F., Guedes De Oliveira P., Principe J.C. (1987) A study on the best order for autoregressive EEG modelling. International Journal of Bio-Medical Computing. 20(1-2): 41-50.

[20] Franaszczuk, P.J., Blinowska, K.J., & Kowalczyk, M. (1985). The application of parametric multichannel spectral estimates in the study of electrical brain activity. Biol Cybern., 51:239–47.

[21] Erla, S., Faes, L., Tranquillini, E., Orrico, D., & Nollo, G. (2009) Multivariate autoregressive model with instantaneous effects to improve brain connectivity estimation. Int J Bioelectromagn 11(2):74–79

[22] Vasios, C.E., Matsopoulos, G.K., Nikita, K.S., & Uzunoglu, N. (2003). Classification of event-related potentials using multivariate autoregressive modeling combined with simulated annealing. Journal of Automatic Control, University of Belgrade, 13(1), 7-11.

[23] Anderson, C.W., Stolz, E.A., & Shamsunder, S. (1998). Multivariate autoregressive models for classification of spontaneous electroencephalographic signals during mental tasks. IEEE Trans Biomed Eng, 45 (3), 277-286.

[24] Orekhova EV, Stroganova TA, Posikera IN. (1999). Theta synchronisation during sustained anticipatory attention in infants over the second half of the first year of life. International Journal of Psychophysiology. 32:151–172.

[25] Faes, L., & Nollo, G. (2011). Multivariate frequency domain analysis of causal interactions in physiological time series. In Biomedical Engineering, Trends in Electronics, Communications and Software; AN Laskovski (ed); INTECH, pp. 403-428. ISBN: 978-953-307-475-7

[26] Delorme, A., & Makeig, S. (2004). EEGLAB: an open source toolbox for analysis of single-trial EEG dynamics including independent component analysis. Journal of neuroscience methods, 134(1), 9-21.

[27] Lachaux, J. P., Rodriguez, E., Martinerie, J., & Varela, F. J. (1999). Measuring phase synchrony in brain signals. Human brain mapping, 8(4), 194-208.

[28] Mallat, S. (1999). A wavelet tour of signal processing. Academic press.

[29] Dunnett, C. W. (1955). A multiple comparison procedure for comparing several treatments with a control. Journal of the American Statistical Association, 50(272), 1096-1121.

[30] Hothorn, L.A. (2016). The two-step approach – a significant ANOVA F-test before Dunnett's comparisons against a control – is not recommended. Communications in statistics – Theory and methods. 45 (11), 3332-3343.

[31] Howell, D.C. (2010). Statistical methods for psychology (7th ed). Belmont, CA: Cengage Wadsworth.