SCIENTIFIC REPORTS

OPEN

Infant-directed speech facilitates seven-month-old infants' cortical tracking of speech

Marina Kalashnikova[1], Varghese Peter[1], Giovanni M. Di Liberto[2,3], Edmund C. Lalor[2,4] & Denis Burnham[1]

This study assessed cortical tracking of temporal information in incoming natural speech in seven-month-old infants. Cortical tracking refers to the process by which neural activity follows the dynamic patterns of the speech input. In adults, it has been shown to involve attentional mechanisms and to facilitate effective speech encoding. However, in infants, cortical tracking or its effects on speech processing have not been investigated. This study measured cortical tracking of speech in infants and, given the involvement of attentional mechanisms in this process, cortical tracking of both infant-directed speech (IDS), which is highly attractive to infants, and the less captivating adult-directed speech (ADS), were compared. IDS is the speech register parents use when addressing young infants. In comparison to ADS, it is characterised by several acoustic qualities that capture infants' attention to linguistic input and assist language learning. Seven-month-old infants' cortical responses were recorded via electroencephalography as they listened to IDS or ADS recordings. Results showed stronger low-frequency cortical tracking of the speech envelope in IDS than in ADS. This suggests that IDS has a privileged status in facilitating successful cortical tracking of incoming speech which may, in turn, augment infants' early speech processing and even later language development.

Cortical tracking refers to the process by which cortical activity tracks dynamic patterns of incoming information, in this case speech input. This applies to both low-level spectrotemporal speech features as well as higher-level speech-specific information[1–5]. The speech envelope contains linguistic information across multiple time scales: at the phonological rate (about 30–50 Hz, corresponding to the gamma band of neural oscillations at which information such as place of articulation, e.g., /b/ vs. /d/ and voicing, /p/ vs. /b/ is conveyed); at the syllabic rate (about 4–8 Hz, corresponding to the theta band of neural oscillations); and at the lexical and phrasal rate (<2 Hz, within the delta band of neural oscillations). Simultaneous cortical tracking of the speech signal across these different temporal scales allows sampling of the incoming speech stream during speech processing[1,6].

Neurophysiological indices of cortical tracking of natural speech can be successfully extracted from electroencephalography (EEG) or magnetoencephalography (MEG) recordings in adult and child participants. Importantly, accurate tracking of the speech envelope has been shown to be affected by changes in intelligibility[7,8] and by mechanisms such as multisensory integration[9]. In addition, selective attention[6,10–12] has been shown to play a significant role in facilitating cortical tracking as demonstrated when adults and school-aged children are presented with different strings of speech input to each ear and asked to switch their attention from one string to the other[3,6,10–12]. In such studies, recordings of neural activity at the theta band (syllabic rate) using MEG show that neural activity is correlated with the acoustic amplitude envelope of both the attended and unattended streams, but that the patterns of correlation differ for both streams as a function of attention, with more accurate tracking recorded in response to the attended stream.

These findings provide strong evidence that cortical tracking involves attentional mechanisms, enhancing listeners' ability to focus selectively on a single speech stream and to filter out other competing auditory information (known as the 'cocktail party' effect). However, these studies have focused solely on adults and school-aged

[1]The MARCS Institute for Brain, Behaviour and Development, Western Sydney University, Locked Bag 1797, Penrith, 2527, Australia. [2]School of Engineering, Trinity Centre for Bioengineering, and Trinity College Institute of Neuroscience, Trinity College Dublin, Dublin, Ireland. [3]Laboratoire des Systèmes Perceptifs, Ecole Normale Supérieure, Paris, 75005, France. [4]Department of Biomedical Engineering and Department of Neuroscience, University of Rochester, Rochester, New York, 14627, USA. Correspondence and requests for materials should be addressed to M.K. (email: m.kalashnikova@westernsydney.edu.au)

children – proficient language users whose extensive phonetic, syntactic, and semantic linguistic competence facilitates processes of parsing and encoding natural speech; and who can be asked to direct their attention to a particular speech stream from several competing speech streams. This is not the case for young infants. Therefore, we aim to tackle two core questions in this study: whether cortical tracking of speech can be measured in young preverbal infants, and if so, whether it is responsive to or even facilitated by augmented attention-grabbing speech. The first question is relevant from a methodological perspective as we propose to evaluate the validity of a measure with infants that has previously been used only with older participants. Furthermore, the answer to this first question is theoretically important, for if there is such cortical tracking of speech, then it would mean that preverbal infants' endogenous neural oscillations entrain to incoming speech before any extensive knowledge of the phonetic, semantic, and syntactic properties of their native language is acquired. The second question is also theoretically important: if infants' cortical tracking is affected by the attention-grabbing properties of speech input, then it would imply that infant directed speech, the type of speech input to which infants are exposed in daily interactions with their parents, may facilitate processes of cortical tracking of speech and early speech processing.

While evidence of cortical tracking processes early in life is limited, Telkemeyer and colleagues[13] have shown that the newborn brain is already sensitive to the temporal structure of speech. They presented newborns with frequency-modulated non-speech stimuli corresponding to the phonological and the lexical speech rates and recorded their electrophysiological and haemodynamic responses that were considered to be equivalent to the adult auditory steady-state response (ASSR), which is phase-locked to the amplitude envelope of auditory input. Infants' neural responses were shown to be tuned to these non-speech modulations, and tuned similarly for the analogues of the phonemic and lexical speech rates. However, to date, cortical tracking of *continuous natural speech* has not been assessed in infants. This is the first aim of this study.

The second aim of this study is to investigate whether infants' cortical tracking is facilitated by what we know to be attentionally-salient to infants, infant-directed speech. Young infants are exposed to extensive linguistic input, both directed to them, infant-directed speech (IDS), and directed to adults around them, adult-directed speech (ADS). These two speech registers differ markedly: compared to ADS, IDS is characterised by slower tempo and speech rate[14], regularised rhythm[15–17], higher emotional content[18], higher pitch and greater pitch range[19], simplified grammatical structure[20], and acoustic exaggeration of speech sounds[21–23]. Given its greater capacity to capture infants' attention, IDS may afford better opportunities for cortical tracking of incoming speech in young infants than might ADS. There are two lines of evidence that provide traction for this possibility.

First, infants prefer IDS to ADS (see[24] for a review). This preference is present in newborns[25] even when IDS is produced by unfamiliar females or males[19] or in a foreign language[26]. Thus, it appears that preference for IDS is driven by its general acoustic and prosodic qualities rather than any particular indexical characteristics. This is further evidenced by neurophysiological studies showing greater cortical activity in temporal and frontal sites for infants up to 12 months of age in response to naturally-produced IDS compared to ADS (functional near-infrared spectroscopy[27,28]; electroencephalography[29]).

Second, exposure to IDS appears to facilitate linguistic processing during the child's first years of life. Behavioural and neurophysiological studies indicate that IDS is better than ADS in promoting performance on a variety of linguistic tasks such as speech sound discrimination[30,31], familiar word recognition[32,33], and word learning[34,35]. Thus, it is possible that by using IDS, parents unconsciously produce not only the type of speech that their infants attend to and prefer, but also the type of speech that assists their infants in the challenging task of learning their native language[36].

While the mechanisms by which IDS might facilitate early language development remain unspecified[37], it may just be that IDS facilitates the cortical tracking of speech, which in turn might underlie enhanced performance in the abovementioned linguistic tasks. In this study, we investigate (i) pre-verbal infants' processing of continuous speech by measuring the tracking of incoming speech by cortical activity; and (ii) whether IDS has a privileged role in facilitating the process of cortical tracking compared to ADS. Our recent research provides a framework for investigating cortical tracking of continuous speech features, e.g., amplitude envelope and phonetic distinctions, using non-invasive EEG[2,38,39]. The approach is based on a ridge regression fit between the speech envelope and the EEG signal and allows prediction of the resulting EEG (forward-modelling), the reconstruction of the speech envelope (backward-modelling), and the derivation of quantitative measures that have been linked to cortical entrainment[9,12,40]. The regression model weights (temporal response functions, TRFs) can be studied in terms of their spatio-temporal dynamics in a similar manner to event related potentials (ERPs[9]), with the important difference that continuous natural stimuli can be used. In particular, this includes examining the distribution of the TRF weights across the scalp at different latencies, that is at different relative time lags between the ongoing speech signal and the ongoing EEG signal. For example, a latency of 100 msec means the impact that a change in the stimulus at time t has on the EEG at time $t + 100$ msec.

This is the first study to use this method with preverbal infants. Here, the level of coupling between infants' cortical activity and the speech envelope in naturally-produced IDS vs. ADS is indexed by relative TRF fit and the relative accuracy of prediction of the unseen EEG signal based on the envelope of the speech signal. There are two specific predictions. First, to confirm that IDS is more attentionally-salient than ADS[25], we expected greater frontal EEG power to IDS than to ADS measured in cortical responses at the theta band[29]. Second, we predicted that, similar to adult and school-age children, infants' cortical tracking of the speech stimuli will be indexed by a correlation between their cortical responses and the low-level acoustic envelope of the speech signal. However, given the differences in attentional salience between IDS and ADS, we expected the envelope to be more strongly reflected in the EEG responses to IDS than ADS.
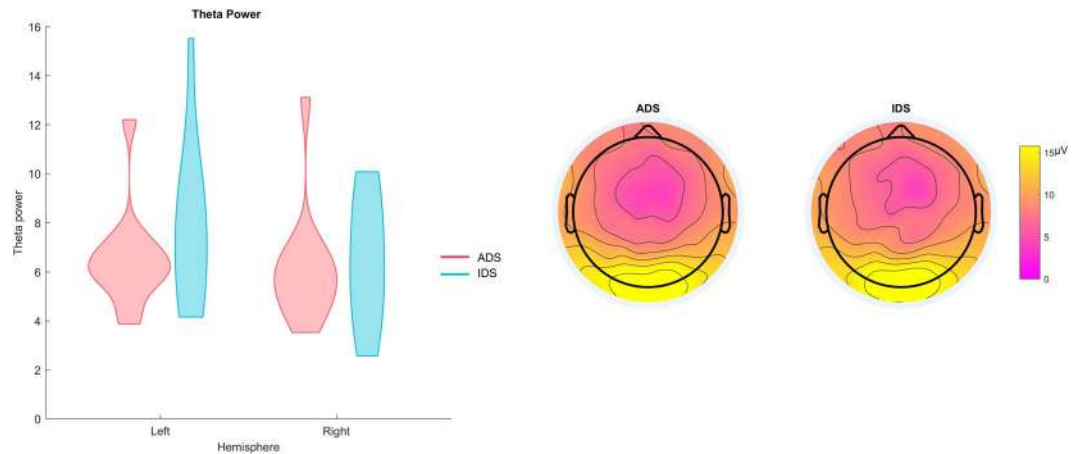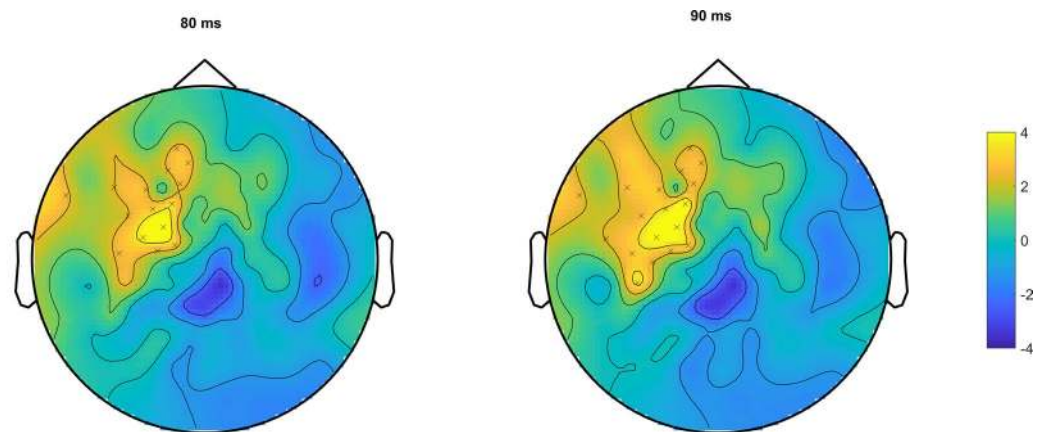
Figure 1. Results of the EEG power analysis averaged for all epochs recorded in IDS and ADS (the left panel displays the EEG power distribution in the theta-band range (4–8 Hz) for IDS and ADS across hemispheres and the right panel displays the scalp topography of the Mean theta power).

X = statistically signficant cluster of t-scores (*N* = 12, *p* <.05)

Figure 2. Results of cluster permutation statistics on the TRFs for IDS and ADS registers.

Results

EEG Power Analysis. Amplitude of the EEG power distribution in the theta-band range (4–8 Hz) across the left and right hemispheres is shown in Fig. 1. A two-way Analysis of Variance (ANOVAs) with register (ADS, IDS) and hemisphere (left, right) as the within-subjects factors and theta power as the dependent variable showed a main effect of hemisphere, $F(1,11) = 10.15$, $p = 0.001$, $\eta^2 = 0.48$, indicating that the theta power over the left hemisphere ($M = 7.14$, $SE = 0.73$) was significantly larger than over the right hemisphere ($M = 6.28$, $SE = 0.64$). Neither the main effect of register, $F(1,11) = 2.28$, $p = 0.16$, $\eta^2 = 0.17$, nor the interaction between register and hemisphere, $F(1,11) = 2.26$, $p = 0.16$, $\eta^2 = 0.16$, were significant. Therefore, contrary to our prediction, theta power did not differ between the IDS and ADS conditions, only between hemispheres, irrespective of register.

TRF Analysis. In order to detect any difference between TRFs for the IDS and ADS registers as a function of speech-EEG time lag, we used a cluster-based permutation analysis[40]. This allows for the identification of clusters of electrodes in which significant response differences between the two registers are detected while controlling for Type I error that may arise due to multiple tests conducted for each electrode (see Method section). The results of this analysis revealed significant differences between the TRFs for IDS and ADS registers in the 80–90 msec time range. Over the left hemisphere, IDS had a significant positive response between 80–90 msec as compared to ADS (Cluster $p = 0.046$). The location of this cluster for each time point is indicated in Fig. 2.

Cortical Entrainment Analysis. The TRFs obtained for ADS and IDS were used to predict the EEG signal using leave-one-out cross-validation[2,38]. Specifically, Pearson's correlation values between the recorded EEG and its prediction were used to index sensor-space cortical tracking of the speech envelope at the individual subject-level. This results in a distribution of correlation values, one for each subject and electrode, that is used to identify significant clusters of scalp electrodes that consistently track the speech envelope. Here, as depicted in
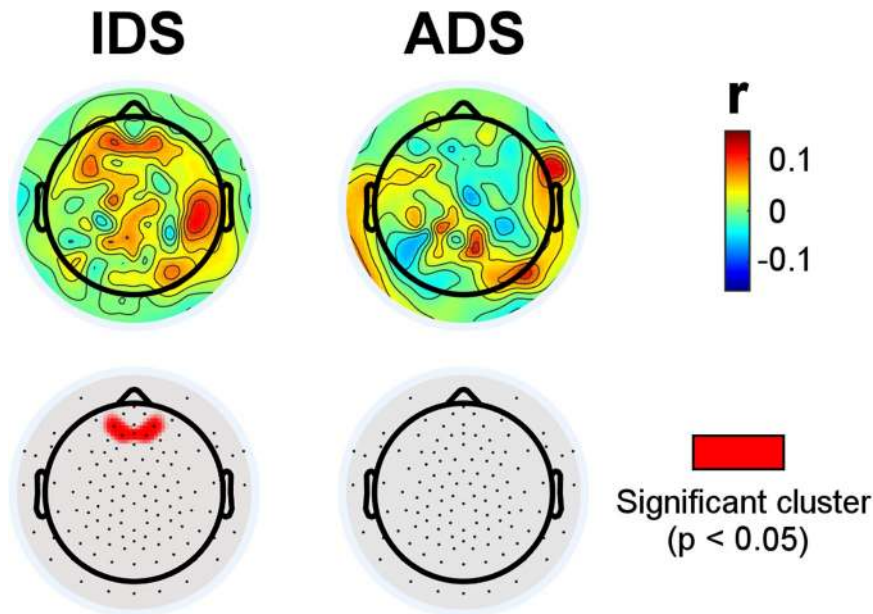
Figure 3. EEG prediction correlations for the IDS and ADS conditions (the top panel shows the correlation *rho* values and the bottom panel highlights the electrodes for which the *rho* values are significantly different from zero).

Fig. 3, significant prediction correlations emerged in response to IDS, forming a cluster composed of 8 electrodes in the frontal area ($p = 0.023$), but this frontal cluster did not emerge in response to ADS. Note that both IDS and ADS showed clusters of electrodes on the right hemisphere that were trending toward statistical significance ($p \simeq 0.1$) but were excluded after the correction for multiple comparisons.

Discussion

This study provides the first evidence for neural tracking of naturally-produced continuous speech by preverbal infants. Neural tracking of continuous speech is associated with effective speech perception[4]. Accordingly, the facility to investigate such mechanisms non-invasively adds a powerful tool to the study of continuous natural speech[3], complementing the traditional event-related approach which is typically used to measure cortical responses to isolated sounds (e.g., syllables, words). In this sense, the TRF approach is an effective method for quantifying cortical tracking of speech sounds using non-invasive EEG[9], and one that is sensitive to the effects of selective attention and multisensory integration[9,38]. Our findings demonstrate that when seven-month-old infants listened to IDS and ADS, there were no significant differences between the theta band power distribution in response to the two registers, but there were significant correlations between infants' cortical activity patterns and the envelope of the speech signal for IDS but not for ADS.

The TRFs were generally larger over the left hemisphere, and they were especially so for IDS than for ADS. This is interesting as the speech envelope was filtered between 1–8 Hz, and according to the asymmetric sampling in time (AST) theory[41], left and right auditory cortices show oscillations at different preferred rates: gamma (25–45 Hz) in the left hemisphere and delta-theta (1–7 Hz) in the right. This left-right asymmetry has been observed regardless of the nature of the stimuli (speech or non-speech) and regardless of the involvement of higher-level speech processing[42,43]. Despite this fact that the left hemisphere is associated with rapid temporal processing[44], our finding aligns with previous studies showing that low frequency information is processed in the left auditory cortex in young pre-verbal infants[45,46]. These findings suggest that the left hemisphere plays a major role in the early processing of low frequency amplitude envelopes. Specifically, increases in left- and front-localised theta oscillations have been proposed to represent the encoding of segmental information in speech[46] as well as greater general attention to speech stimuli in young infants[47].

The human perceptual system is exposed to multiple sources of information at any given time, and the ability to entrain selectively to one speech stream may be crucial in real-time decoding of information at the three levels of the speech processing (e.g., phoneme, syllable, lexical). This may be even more complex in young infants whose native language competence is still being acquired. The findings here indicate that such acquisition may be facilitated by IDS. That is, IDS is an enriched speech register that augments infants' encoding and decoding of speech.

Previous research has shown that infants' performance on a number of language processing tasks is enhanced when stimuli are presented in IDS compared to ADS. The results of this study add neural tracking of natural speech to the list of processes that are facilitated by speech input via IDS. Nevertheless, it is unclear from these findings what specific characteristics of IDS lead to the processing benefits attributed to this register. Two possibilities suggest themselves. First, as stated in the predictions of this study, these findings may be due to top-down processes such as greater attention elicited by the prosodic properties of IDS. This conclusion dovetails with previous research with adult participants in attention-based paradigms[40], and the evidence that IDS attracts greater

attentional responses in infants[25,27–29]. This explanation can be further strengthened by the fact that our analyses compared neural tracking of two speech signals, IDS and ADS, instead of a silence baseline[29] or a non-speech condition, e.g., amplitude modulated noise[13], indicating that the general attention-grabbing qualities of IDS rather than specific linguistic information underlie the difference between the two registers uncovered here.

The second possibility is that bottom-up processes play a role in facilitating cortical tracking in IDS, as IDS is also distinguished from ADS on characteristics such as phonetic realisation[21], greater pitch range[48], and regularised speech rhythm[16]. This view has been proposed in previous literature whereby specific components of IDS have been identified in relation to individual linguistic tasks. For instance, exaggerated productions of phonemes in IDS have been related to better discrimination of native speech sounds[49], exaggerated prosodic patterns have been proposed to facilitate continuous speech segmentation[50], and the vowel hyperarticulation and speech rate components of IDS have been shown to facilitate lexical processing[33]. The two possibilities proposed here are of course not mutually exclusive. It is plausible that all these prosodic and linguistic components of IDS act in unison: greater correlation between the stimulus amplitude envelope and the neural activity envelope recorded with EEG may be a product of infants' tendency to direct their attention to IDS, and also of the encoding of individual features of the incoming speech string.

Tracking of the acoustic envelope of incoming speech by the endogenous neural oscillations has been linked to successful encoding of auditory speech. This occurs when adult listeners selectively attend to the speech input in their environment[40]. This study demonstrates that a similar process occurs when infants as young as seven months of age listen to speech. Most importantly, there is a greater correlation between neural activity and the speech envelope when infants listen to IDS compared to ADS. These findings suggest that the special register that parents spontaneously use when addressing their young infants (and which has been found to foster early linguistic processing) facilitates speech encoding at its multiple timescales even at the earliest stages of language acquisition. Further work is required to locate the precise source of this early difference in processing IDS and ADS input as well as to identify the developmental time course for neural tracking of ADS.

Method

Participants. Twelve seven-month-old infants participated in the experiment (6 female; M age = 225.1 days, SD = 9.1). All infants were acquiring English as their first language, were born full-term, and were not at-risk for cognitive or language delay. Seven additional infants were tested, but their data were removed from the analysis as five infants had more than 20 bad channels and two infants did not complete the experiment. This study was approved by the Human Research Ethics Committee at Western Sydney University (approval number 9142). Prior to the study, the primary caregiver of each infant completed an informed consent form, and they were informed that the procedure would be immediately discontinued if they wished so, or if their infant showed any signs of distress. This study followed the approved protocol regarding participant recruitment, data collection, and data management.

Stimuli. The stimuli consisted of recordings of naturally-produced IDS and ADS. These recordings were produced by a female Australian English speaker when she interacted with her seven-month-old infant (IDS) or an experimenter (ADS). During the IDS recording, the speaker and her infant sat alone in an infant laboratory room, and she was instructed to interact naturally with her baby. She was provided with soft toys and pictures to facilitate the interaction. During the ADS recording, the speaker was interviewed by a female experimenter, also a native speaker of Australian English, in the same laboratory room and she was asked to comment about the IDS session. The infant was not present during the ADS recording. A head-mounted microphone (AudioTechnica AT892) connected to Adobe Audition CS6 software via an audio input/output device (MOTU Ultralite MK3) was used during the speech recordings.

The ADS recording was 481 seconds in duration and the IDS recording was 486 seconds in duration. As expected, compared to ADS, IDS had higher pitch and greater pitch range (IDS: M F0 = 202.19 Hz, F0 range = 385.88 Hz; ADS: M F0 = 171.53 Hz, F0 range = 298.59 Hz), slower speech rate (IDS: 2.85 words/second, 1384 words in total; ADS: 3.76 words/second, 1810 words in total), and hyperarticulated vowels (based on the area of the triangle resulting from plotting F1 and F2 values for the three corner vowels /i/, /u/, and /a/; IDS area = 17263.26 Hz^2, ADS area = 9455.50 Hz^2).

EEG Recording. The infants sat on their parent's lap approximately 1 m from an LCD screen. Stimuli were presented through audio speakers at 75 dB SPL. In order to maintain infants' attention, a coloured checkerboard was presented on the screen which changed colour every 30 seconds. While infants listened to the speech stimuli, their continuous EEG was recorded using a 129 channel Hydrocel Geodesic Sensor Net (HCGSN), NetAmps 300 amplifier and NetStation 4.5.7 software (EGI Inc) at a sampling rate of 1000 Hz with the reference electrode placed at Cz. The electrode impedances were kept below 50 kΩ. The continuous EEG was saved for offline analysis.

EEG Pre-processing. The EEG analysis was performed using EEGLAB[51], FieldTrip[52], the mTRF toolbox[38], and custom scripts in MATLAB2014a[53]. Since the infant EEG recordings are noisy due to infant movements, we applied artifact subspace reconstruction (ASR[54]) to remove noise. ASR uses a sliding window technique whereby each window of EEG data is decomposed via principal component analysis so it can be compared statistically with data from a calibration dataset. Within each sliding window the ASR algorithm identifies principal subspaces, which significantly deviate from the baseline EEG and then reconstructs these subspaces using a mixing matrix computed from the baseline EEG recording. In this study, we used a sliding window of 500 msec and a threshold of 20 standard deviations to identify corrupted subspaces. The noisy channels that were removed during the ASR procedure were later replaced by averaging the neighbouring channels weighted by distance. The EEG was then analysed in two ways: EEG power analysis and temporal response function (TRF) analysis.
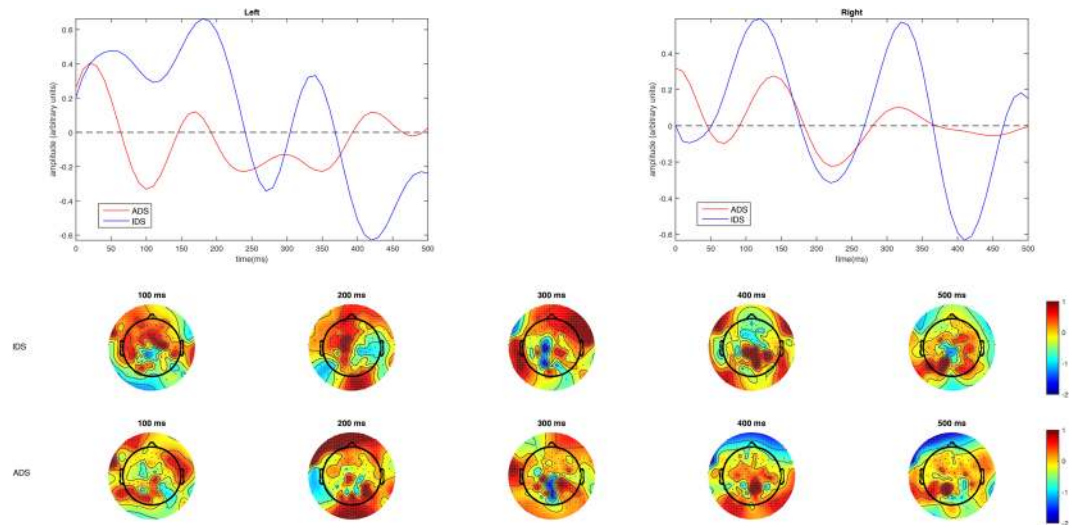
Figure 4. Temporal Response Function (TRF) analysis for IDS and ADS registers (the top panel displays the TRFs for ADS and IDS over the left and right hemispheres, and the bottom panel displays the scalp topography of the TRFs from 100 to 500 msec).

EEG Power Analysis. First EEG data were down-sampled to 250 Hz for computational reasons. The EEG was then divided into one second non-overlapping epochs starting from the onset of the ADS and IDS recordings. EEG epochs with amplitude fluctuations exceeding $\pm 100\,\mu V$ were removed. The data were then re-referenced to the average of all the electrodes. All participants had at least 100 artifact-free epochs (corresponding to 100 seconds of EEG data; ADS $M = 206.5$, $SD = 25.23$; IDS $M = 203.52$, $SD = 31.55$). All artifact-free EEG data were analysed using a discrete Fourier transform (DFT), with a Hanning window of one second width and 50% overlap. Power (μV^2) was derived from the DFT output in the 4–8 Hz frequency band for ADS and IDS conditions.

EEG power in the theta region was then computed across two scalp locations: frontal left and frontal right. Frontal electrodes were selected for the analysis as auditory stimuli generate larger responses at frontal electrodes (see the Supplementary Fig. S1 for a graphical representation of these electrode groupings).

TRF Analysis. The data were first re-referenced to average reference. The amplitude envelope of speech between 1–8 Hz was then extracted using Hilbert transform. In order to reduce the computational time, both speech envelope and EEG were down-sampled to a sampling rate of 128 Hz. Twenty-one electrodes on the periphery of the electrode net were removed as it generally records noisy data in infants, and it is common to remove these peripheral electrodes when conducting the analysis of infant EEG[55,56].

A ridge regression model was fitted from the speech envelope to the EEG signal for every participant and channel[38]. The window was restricted to the 0 msec to 500 msec window because no visible response emerged out of that lag interval. The regularisation parameter of the model (lambda) was chosen using a quantitative procedure that aims at producing an optimal model fit. A lambda of 1 was selected for the current analysis. The resulting regression weights, referred to as temporal response functions (TRFs), were studied in terms of their spatial and temporal dynamics, similar to an ERP analysis. The TRFs obtained from individual participants were averaged to obtain the grand averaged TRF waveform.

Five-fold leave-one-out cross-validation was used to assess how well the unseen EEG data could be predicted. This was achieved by using the TRF fit on 4 folds to predict the EEG data of the 5th fold, and to iterate this procedure for all combinations. If EEG can be predicted with accuracy significantly greater than zero, it can be asserted that the EEG is reflecting the encoding of the envelope of speech[38]. Prediction accuracy was measured by calculating Pearson's (r) linear correlation coefficient between the predicted and original EEG responses at each electrode channel. The time window that best captures the stimulus-response mapping is used for EEG prediction (i.e., T_{min}, T_{max}). This is identified by examining the TRFs on a broad time window (e.g., -100 msec to 600 msec) and then choosing the temporal region of the TRF that includes all relevant components that map the stimulus to the EEG with no evident response outside of this range. In this case, the time window chosen for this quantitative analysis was 0–500 msec. Figure 4 shows the TRFs as well as the scalp topography of the TRF peaks for ADS and IDS for this time window.

Cluster Permutation Analysis. In order to assess the difference between TRFs for the IDS and ADS conditions at any time point of the recording, a cluster-based permutation analysis was employed[57]. In this analysis, multiple t-tests are computed at every electrode and every time point. From this analysis, clusters of electrodes and time points in which the response significantly differs between conditions are identified. These clusters are then formed over space by grouping electrodes that have significant initial t-test values at the same time point. The sum of all t-scores within each cluster provides a cluster-level t-score (mass t-score). A permutation approach is then used to control for Type I errors, by randomly assigning conditions and repeating the multiple t-tests (1000 iterations) in order to build a data-driven null hypothesis distribution. The relative location of each observed real

cluster mass *t*-score within the null hypothesis distribution indicates how probable such a score would be if the null hypothesis were true. The significance of a cluster is determined by whether it falls in the highest or the lowest 2.5th percentile of the corresponding distribution.

All data and tools used for analyses in this study are available upon request to the first author.

References

1. Ding, N., Melloni, L., Zhang, H., Tian, X. & Poeppel, D. Cortical tracking of hierarchical linguistic structures in connected speech. *Nature Neuroscience* **19**(1), 158–164 (2016).
2. Di Liberto, G. M., O'Sullivan, J. A. & Lalor, E. C. Low-frequency cortical entrainment to speech reflects phoneme-level processing. *Current Biol.* **25**(19), 2457–2465 (2015).
3. Ding, N. & Simon, J. Z. Emergence of neural encoding of auditory objects while listening to competing speakers. *PNAS* **109**(29), 11854–11859 (2012).
4. Ghitza, O., Giraud, A. L. & Poeppel, D. Neuronal oscillations and speech perception: critical-band temporal envelopes are the essence. *Frontiers in Human Neuroscience* **6**, 340 (2012).
5. Leong, V. & Goswami, U. Acoustic-emergent phonology in the amplitude envelope of child-directed speech. *PLoS One* **10**(12), e0144411 (2015).
6. Power, A. J., Mead, N., Barnes, L. & Goswami, U. Neural entrainment to rhythmically presented auditory, visual, and audio-visual speech in children. *Frontiers in Psychology* **3**, 216 (2012).
7. Di Liberto, G. M., Lalor, E. C. & Millman, R. E. Causal cortical dynamics of a predictive enhancement of speech intelligibility. *Neuroimage* **166**(1), 247–258 (2018).
8. Peelle, J. E., Gross, J. & Davis, M. H. Phase-locked responses to speech in human auditory cortex are enhanced during comprehension. *Cerebral Cortex* **23**(6), 1378–1387 (2012).
9. Crosse, M. J., Di Liberto, G. M., Bednar, A. & Lalor, E. C. The multivariate temporal response function (mTRF) toolbox: a MATLAB toolbox for relating neural signals to continuous stimuli. *Frontiers in Human Neuroscience* **10**, 604 (2016).
10. Golumbic, E. M. Z. *et al.* Mechanisms underlying selective neuronal tracking of attended speech at a "cocktail party". *Neuron* **77**(5), 980–991 (2013).
11. Kerlin, J. R., Shahin, A. J. & Miller, L. M. Attentional gain control of ongoing cortical speech representations in a "cocktail party". *J. Neuroscience* **30**(2), 620–628 (2010).
12. O'Sullivan, J. A. *et al.* Attentional selection in a cocktail party environment can be decoded from single-trial EEG. *Cerebral Cortex* **25**(7), 1697–1706 (2015).
13. Telkemeyer, S. *et al.* Sensitivity of newborn auditory cortex to the temporal structure of sounds. *J. Neuroscience* **29**(47), 14726–14733 (2009).
14. Panneton, R., Kitamura, C., Mattock, K. & Burnham, D. Slow speech enhances younger but not older infants' perception of vocal emotion. *Research in Human Development* **3**(1), 7–19 (2006).
15. Lee, C. S., Kitamura, C., Burnham, D. & Todd, N. P. On the rhythm of infant- versus adult-directed speech in Australian English. *JASA* **136**(1), 357–365 (2014).
16. Leong, V., Kalashnikova, M., Burnham, D., & Goswami, U. The temporal modulation structure of infant-directed speech. *Open Mind*, https://doi.org/10.1162/OPMI_a_00008 (2017).
17. Payne, E., Post, B., Astruc, L., Prieto, P. & Vanrell, M. M. Rhythmic modification in child directed speech. *Oxford University Working Papers in Linguistics, Philology & Phonetics* **12**, 123–144 (2009).
18. Kitamura, C. & Burnham, D. Pitch and communicative intent in mother's speech: Adjustments for age and sex in the first year. *Infancy* **4**(1), 85–110 (2003).
19. Fernald, A. *et al.* A cross-language study of prosodic modifications in mothers' and fathers' speech to preverbal infants. *JCL* **16**, 477–501 (1989).
20. Soderstrom, M. Beyond babytalk: Re-evaluating the nature and content of speech input to preverbal infants. *Developmental Review* **27**(4), 501–532 (2007).
21. Burnham, D., Kitamura, C. & Vollmer-Conna, U. What's new pussycat? On talking to babies and animals. *Science* **296**(5572), 1435–1435 (2002).
22. Kalashnikova, M., Carignan, C. & Burnham, D. The origins of babytalk: Smiling, teaching or social convergence? *RSOS* **4**(8), 170306 (2017).
23. Kuhl, P. *et al.* Cross-language analysis of phonetic units in language addressed to infants. *Science* **277**(5326), 684–686 (1997).
24. Dunst, C., Gorman, E. & Hamby, D. Preference for infant-directed speech in preverbal young children. *Center for Early Literacy Learning* **5**(1), 1–13 (2012).
25. Cooper, R. P. & Aslin, R. N. Preference for infant-directed speech in the first month after birth. *Child Dev.* **61**(5), 1584–1595 (1990).
26. Werker, J. F., Pegg, J. E. & McLeod, P. J. A cross-language investigation of infant preference for infant-directed communication. *Infant Behaviour and Development* **17**, 323–333 (1994).
27. Naoi, N. *et al.* Cerebral responses to infant-directed speech and the effect of talker familiarity. *Neuroimage* **59**(2), 1735–1744 (2012).
28. Saito, Y. *et al.* Frontal cerebral blood flow change associated with infant directed speech. *Archives of Diseases in Childhood* **92**, 113–116 (2007).
29. Santesso, D. L., Schmidt, L. A. & Trainor, L. J. Frontal brain electrical activity (EEG) and heart rate in response to affective infant-directed (ID) speech in 9-month-old infants. *Brain and Cognition* **65**(1), 14–21 (2007).
30. Peter, V., Kalashnikova, M., Santos, A. & Burnham, D. Mature neural responses to infant-directed speech but not adult-directed speech in pre-verbal infants. *Scientific Reports* **6**, 34273 (2016).
31. Zhang, Y. *et al.* Neural coding of formant-exaggerated speech in the infant brain. *Dev. Sci.* **14**(3), 566–581 (2011).
32. Singh, L., Nestor, S., Parikh, C. & Yull, A. Influences of infant-directed speech on early word recognition. *Infancy* **14**(6), 654–666 (2009).
33. Song, J. Y., Demuth, K. & Morgan, J. Effects of the acoustic properties of infant-directed speech on infant word recognition. *JASA* **128**(1), 389–400 (2010).
34. Graf-Estes, K. & Hurley, K. Infant-directed prosody helps infants map sounds to meanings. *Infancy* **18**(5), 797–824 (2013).
35. Ma, W., Golinkoff, R. M., Houston, D. M. & Hirsh-Pasek, K. Word learning in infant- and adult-directed speech. *Language Learning and Development* **7**(3), 185–201 (2011).
36. Kuhl, P. A new view of language acquisition. *Pnas* **97**(22), 11850–11857 (2000).
37. Cristia, A. Input to language: The phonetics and perception of infant-directed speech. *Language and Linguistics Compass* **7**(3), 157–170 (2013).
38. Crosse, M. J., Butler, J. S. & Lalor, E. C. Congruent visual speech enhances cortical entrainment to continuous auditory speech in noise-free conditions. *J. Neuroscience* **35**(42), 14195–14204 (2015).
39. Lalor, E. C., Power, A. J., Reilly, R. B. & Foxe, J. J. Resolving precise temporal processing properties of the auditory system using continuous stimuli. *J. Neurophysiology* **102**(1), 349–359 (2009).
40. Ding, N. & Simon, J. Z. Cortical entrainment to continuous speech: Functional roles and interpretations. *Frontiers in Human Neuroscience* **8**, 311 (2014).

41. Poeppel, D. The analysis of speech in different temporal integration windows: cerebral lateralization as 'asymmetric sampling in time'. *Speech Communication* **41**(1), 245–255 (2003).
42. Giraud, A. L. & Poeppel, D. Cortical oscillations and speech processing: Emerging computational principles and operations. *Nature Neuroscience* **15**(4), 511–517 (2012).
43. Morillon, B. *et al*. Neurophysiological origin of human brain asymmetry for speech and language. *PNAS* **107**(43), 18688–18693 (2010).
44. Zatorre, R. J. & Belin, P. Spectral and temporal processing in human auditory cortex. *Cereb. Cortex* **11**, 946–953 (2001).
45. Minagawa-Kawai, Y. *et al*. Optical brain imaging reveals general auditory and language specific processing in early infant development. *Cereb. Cortex* **21**, 254–261 (2011).
46. Ortiz-Mantilla, S., Hämäläinen, J. A., Musacchia, G. & Benasich, A. A. Enhancement of gamma oscillations indicates preferential processing of native over foreign phonemic contrasts in infants. *J. of Neuroscience* **33**(48), 18746–18754 (2013).
47. Orekhova, E. V., Stroganova, T. A., Posikera, I. N. & Elam, M. EEG theta rhythm in infants and preschool children. *Clin. Neurophysiol.* **117**, 1047–1062 (2006).
48. Trainor, L. J., Austin, C. M. & Desjardins, R. N. Is infant-directed speech prosody a result of the vocal expression of emotion? *Psych. Sci.* **11**(3), 188–195 (2000).
49. Liu, H. M., Kuhl, P. & Tsao, F. M. An association between mothers' speech clarity and infants' speech discrimination skills. *Dev. Sci.* **6**(3), F1–F10 (2003).
50. Thiessen, E., Hill, E. A. & Saffran, J. R. Infant-directed speech facilitates word segmentation. *Infancy* **7**(1), 53–71 (2005).
51. Delorme, A. & Makeig, S. EEGLAB: An open source toolbox for analysis of single-trial EEG dynamics including independent component analysis. *J. Neuroscience Methods* **134**(1), 9–21 (2004).
52. Oostenveld, R., Fries, P., Maris, E., & Schoffelen, J. M. FieldTrip: Open Source Software for Advanced Analysis of MEG, EEG, and Invasive Electrophysiological Data. *Computational Intelligence and Neuroscience* (2011).
53. The Mathworks, Inc. MATLAB R2014a.Natick, Massachusetts: The Mathworks, Inc (2015).
54. Mullen, T. *et al*. Real-time modeling and 3D visualization of source dynamics and connectivity using wearable EEG. In Conference proceedings: Annual International Conference of the IEEE Engineering in Medicine and Biology Society. *IEEE Engineering in Medicine and Biology Society. Conference*. NIH Public Access 2184 (2013).
55. Brusini, P. *et al*. Ambiguous function words do not prevent 18-month-olds from building accurate syntactic category expectations: An ERP study. *Neuropsychologia* **98**, 4–12 (2016).
56. Folland, N. A., Butler, B. E., Payne, J. E. & Trainor, L. J. Cortical representations sensitive to the number of perceived auditory objects emerge between 2 and 4 months of age: Electrophysiological evidence. *J. Cognitive Neuroscience* **27**(5), 1060–7 (2015).
57. Maris, E. & Oostenveld, R. Nonparametric statistical testing of EEG-and MEG-data. *J. Neuroscience Methods* **164**(1), 177–190 (2007).

Acknowledgements

Author Contributions

M.K. developed the study concept. M.K., V.P. and D.B. contributed to the study design. Data collection was conducted by M.K. and V.P. Data analysis was conducted by G.L. and V.P. and M.K., G.L., V.P., D.B. and E.L. contributed to the interpretation of the findings. M.K. drafted the manuscript and all authors provided critical revisions and approved the final version of the manuscript.

Additional Information

Supplementary information accompanies this paper at https://doi.org/10.1038/s41598-018-32150-6.

Competing Interests: The authors declare no competing interests.

Publisher's note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.