Cortical tracking of visual rhythmic speech by 5- and 8-month-old infants: Individual differences in phase angle relate to language outcomes up to 2 years

Áine Ní Choisdealbha, Adam Attaheri, Sinead Rocha, Natasha Mead, Helen Olawole-Scott, Maria Alfaro e Oliveira, Carmel Brough, Perrine Brusini, Samuel Gibbon, Panagiotis Boutris, Christina Grey, Isabel Williams, Sheila Flanagan, and Usha Goswami *

Centre for Neuroscience in Education, Department of Psychology, University of Cambridge

*Corresponding author: Usha Goswami, ucg10@cam.ac.uk

Running Title: ***Visual Phase Angle and Language Outcomes***

Acknowledgements:

We are grateful to the families in the Cambridge BabyRhythm UK project for their participation. This work was funded by the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme (grant agreement No.694786).

Data availability statement: Scripts and processed data used for analysis will be made available at https://osf.io/tmuqy/ upon acceptance

This work was reviewed by the Psychology Research Ethics Committee of the University of Cambridge.

The authors declare that they have no competing interests.

Highlights

- Infant preferred phase to visual rhythmic speech predicts language outcomes
- Significant cortical tracking of visual speech is present at 5 and 8 months
- Phase angle to visual speech at 8 months predicted greater receptive and productive vocabulary at 24 months.

Abstract

It is known that the rhythms of speech are visible on the face, accurately mirroring changes in the vocal tract. These low-frequency visual temporal movements are tightly correlated with speech output, and both visual speech (for example, mouth motion) and the acoustic speech amplitude envelope entrain neural oscillations. Low-frequency visual temporal information ('visual prosody') is known from behavioural studies to be perceived by infants, but oscillatory studies are currently lacking. Here we measure cortical tracking of low-frequency visual temporal information by five- and eight-month-old infants using a rhythmic speech paradigm (repetition of the syllable "ta" at 2 Hz). Eye-tracking data was collected simultaneously with EEG, enabling computation of cortical tracking and phase angle during visual-only speech presentation. Significantly higher power at the stimulus frequency indicated that cortical tracking occurred across both ages. Further, individual differences in preferred phase to visual speech related to subsequent measures of language acquisition. The difference in phase between visual-only speech and the same speech presented as auditory-visual at 6- and 9-months was also examined. These neural data suggest that individual differences in early language acquisition may be related to the phase of entrainment to visual rhythmic input in infancy.

Keywords: infant, EEG, visual speech, language acquisition

1. Introduction

Visual speech inputs are known to enhance auditory speech information for adults, particularly in noisy or degraded conditions (e.g., Sumby & Pollack, 1954; Grant & Braida, 1991; Grant & Seitz, 2000). Visual speech information may also enhance speech processing for infants, as already demonstrated behaviourally regarding phonetic information, which can be easy to see on the face (Kuhl & Meltzoff, 1982; Pons et al., 2009; Walden et al., 1977). The theoretical focus adopted here concerns speech prosody and speech rhythm rather than phonetic information, as the acoustic statistics of infant-directed speech (IDS) foreground speech rhythm patterns (Leong et al., 2017). Speech rhythm is also visible on the face (Munhall, Kross & Vatikiotis-Bateson, 2002; Kitamura, Guellaï & Kim, 2014). Indeed, the temporal correspondence between visual and auditory prosody is highlighted by the natural statistics of audiovisual (AV) speech, as lip, jaw, cheek and head movements convey crucial information about the speech amplitude envelope (AE), which carries acoustic prosodic information (Chandrasekaran, Trubanova, Stillittano, Caplier & Ghazanfar, 2009). Low-frequency spatial and temporal modulations in the 2 – 7 Hz range related to mouth opening and closing are of particular importance for AE tracking, causing Chandrasekeran et al. (2009) to observe that the natural temporal features of AV speech signals are optimally structured for the neural (oscillatory) rhythms of the brains of their receivers. In the current report, we investigate whether these oscillatory rhythms are encoding visual prosody in the pre-verbal infant brain, and whether individual differences in the tracking of visual speech have implications for later language acquisition. Speech prosody and speech rhythm play central roles in infant language acquisition (Mehler et al., 1988), but until recently similarities between neural speech processing mechanisms in infants and adults have been relatively under-explored (Attaheri et al., 2022a,b; Tan et al., 2022).

There are many adult studies demonstrating oscillatory cortical tracking of the speech AE for both auditory-only and AV speech (see Giraud & Poeppel, 2012, for review). Consistent with the view that *visual speech* plays an important role in this AE tracking, Bourguignon and colleagues (2020) showed that for adults listening passively to AV natural speech, neural entrainment to the speech AE and to mouth opening/lip movements was tightly coupled (see also Park et al., 2016). Further, when visual-only speech was presented (via silent videos), there were significant low-frequency neural responses in auditory cortex, similar to the responses recorded when speech was heard as well as seen. Bourguignon and colleagues (2020) concluded that auditory neural oscillatory entrainment to prosodic acoustic information in the speech AE occurs even when viewing visual-only speech information. Recent oscillatory studies of speech processing by infants have demonstrated robust low-frequency cortical tracking of the speech AE from birth throughout the first year (Kalashnikova et al., 2019; Ortiz Barajas et al., 2021; Attaheri et al., 2022a; Tan et al., 2022), in response to both auditory-only and AV inputs. Whether robust cortical tracking occurs during visual-only speech processing by infants of multiple ages is currently unknown. Many of the neural mechanisms for speech processing first reported in the adult literature appear to be recruited by the infant brain (e.g., theta-gamma phase amplitude coupling, Attaheri et al., 2022b). Accordingly, other mechanisms found in adults, such as entrainment to visual speech, may be an intrinsic part of the neural speech processing architecture. Visual articulatory cues can precede vocalisations by 200ms or more, and visual gestures such as mouth opening may reset auditory cortex to the "optimal state" for processing the succeeding vocalisations (Schroeder et al., 2008). Such mechanisms might thus be part of the infant toolkit for language learning, and these automatic neural mechanisms may help to scaffold language acquisition.

To investigate this possibility, here we chose to examine infant neural processing of visual-only speech using a rhythmic speech paradigm. The current study is part of the

Cambridge UK BabyRhythm study, which tests key tenets of temporal sampling (TS) theory (Goswami, 2011; 2019) in infants using a range of rhythmic measures. TS theory was originally derived from neural studies of children with developmental dyslexia, who present primarily with phonological (speech-sound) processing difficulties. TS theory focused first on the neural processing of speech rhythm in dyslexia, identifying delta-band differences in the phase alignment of low frequency neural oscillations to speech between children with and without developmental dyslexia in rhythmic syllable repetition tasks (Power et al., 2013). At the individual level, the angle of phase alignment in these tasks related to phonological processing and reading ability (Power et al., 2012, 2013). Related dyslexia studies replicated this atypical phase effect for rhythmic syllable repetition (Keshavarzi et al., 2022a), and also indicated less accurate encoding of delta-band speech envelope information in connected speech tasks using an mTRF approach (Power et al., 2016; Keshavarzi et al., 2022b). The Cambridge UK BabyRhythm study thus set out to use similar tasks and measures with typically-developing infants, and has reported on cortical tracking of sung speech using the mTRF (Attaheri et al., 2022a,b) and on phase relations with rhythmic speech using a syllable repetition task in auditory-only and auditory-visual (AV) modalities (Ní Choisdealbha et al., 2022, 2023). Although visual-only rhythmic speech was studied in children with dyslexia, no phase-related differences were found in comparison to typically-developing children (Power et al., 2013). Nevertheless, it is possible that individual differences in phase alignment to visual-only rhythmic speech might affect the early development of language, before infants begin to speak, and we investigate that possibility here.

There is one prior study of infant cortical tracking of visual-only speech, using IDS and the mTRF analysis approach (Tan et al., 2022). Tan and colleagues reported no significant tracking of visual-only IDS by infants aged five months, despite finding significant tracking of auditory-only and AV IDS by the same infants. However, Tan et al. reported that infants who

paid more attention to the *mouth area* during visual-only speech also showed more accurate cortical tracking of the stimulus. This appears to suggest that greater attention to the rhythmic cue of mouth opening facilitates tracking of visual speech by infants, which could indicate the presence of cortical tracking in some participants. It is known that rhythmic facial and head movements convey prosodic information about the speech AE, because head, eyebrow, lip, jaw and cheek movements are systematically related to speech amplitude and fundamental frequency (e.g. Munhall et al., 2002; Munhall et al., 2004). When a person is speaking, the vocal tract is the source of facial movement, thus speech acoustics can be estimated from face motion (and vice versa) with high reliability (Yehia, Rubin & Vatikiotis-Bateson, 1998; Yehia, Kuratate & Vatikiotis-Bateson, 2002). As information about the AE is visible on the face of a speaker, this information should also be available to the pre-verbal brain, making the negative result reported by Tan et al. (2022) surprising. The rhythmic speech paradigm employed here is a simpler input, and may be able to detect significant cortical tracking of visual-only speech even at 5 months of age.

The rhythmic speech paradigm developed for our infants was modelled on the original work by Power et al. (2012), involving repetition of the syllable "ba" at a rate of 500 ms (2 Hz). Note that a consideration of frequency rates is critical regarding cross-modal speech dynamics, as most visual speech information is recovered from the lower spatial and temporal frequencies (below 7 cycles of visual resolution per face, and 6 – 9 frames per second, respectively), and adult studies of the natural statistics of AV speech highlight the importance of low-frequency rhythms between 2 and 7 Hz (Chandrasekaran et al., 2009). For example, when Chandrasekaran and colleagues studied natural spoken sentences in English and French, they analysed the area of mouth opening as a function of time (visual temporal content). They reported a close temporal correspondence between the area of mouth opening and the wideband acoustic envelope, with a distinct rhythm that was between 2 and 7 Hz. Our prior neural data

with the same infants studied here but using AV sung speech has also highlighted the importance of low-frequency modulations in the AE for cortical tracking (Attaheri et al., 2022a). Infants aged four, seven and eleven months show enhanced cortical tracking in the delta (0.5 – 4 Hz) and theta (4 – 8 Hz) bands compared to higher frequencies (8 – 12 Hz). Low-frequency tracking of visual-only speech has yet to be explored longitudinally, and we provide such information here.

If the human auditory speech processing architecture is engaged automatically by visual speech inputs, and given our prior child data with the rhythmic speech paradigm, then pre-verbal infants should also benefit from visual speech when processing human vocalisations. For example, the phase alignment (or phase angle) between a stimulus and the oscillatory neural response to that stimulus may relate to later language acquisition. Adult studies using natural speech suggest that one key mechanism regarding efficient speech processing is the *phase dynamics* of auditory and visual low-frequency neuronal oscillations (Luo, Liu and Poeppel, 2010). In their MEG experiment, when adult participants were presented with congruent auditory and visual information, the *visual* stream modulated *auditory* cortical activity by aligning auditory oscillatory phase to the *optimal phase angle*, so that the expected auditory input arrived during a high excitability state. Whether the brains of pre-verbal infants would show similar cross-modal neural alignment mechanisms is not currently known. However, earlier data from our BabyRhythm cohort has shown that phase alignment of neural oscillations to auditory speech input is indeed active during infant speech processing, with notable individual differences. For auditory-only rhythmic speech (the syllable repetition task), the phase of infants' neural oscillations aligns consistently to the rhythmic input as young as two months of age (Ní Choisdealbha et al., 2023).

Ní Choisdealbha and colleagues further demonstrated group-level convergence towards a common *preferred phase angle* for rhythmic AV speech during the first year of life,

potentially reflecting convergence to an “optimal” phase angle for speech processing as envisaged by Schroeder et al. (2008). In Ní Choisdealbha et al. (2023), the term “phase angle” or “mean phase angle” referred to the angle of a cortical response at a given frequency at the time at which it was observed, averaged over multiple observations. The same terminology is used here. As the visual stimulus depicted an actor silently repeating a syllable every 500ms (i.e. at a 2Hz rate), observations from the EEG time-frequency data were taken every 500ms. Observations were taken across multiple EEG frequency bins for the purpose of statistical comparison, but the sampling rate was always 500ms. "Preferred phase" thus refers to the tendency for neural oscillations to occur around a particular angle (timepoint) relative to a specific event at which it is measured (e.g. a repeated sound), and hence encompasses both the mean phase angle of the observed oscillation and the consistency with which it occurs at or around the same angle each time it is measured. Preferred phase can either be measured at an individual level, by looking at an individual's repeated responses to a stimulus, or at a group level, by examining whether the preferred phases of the individuals in the group fall into a similar alignment. Group preferred phase may thus indicate the optimal alignment between a stimulus and a neural response regarding the tracking of speech rhythm. Consequently, infants whose phase angles are closer to this “preferred” alignment may have better language outcomes. In our prior investigations with auditory-only and AV rhythmic speech, individual differences in infant preferred phase indeed affected subsequent language development. Infant preferred phase to auditory-only and AV speech was related to a range of measures of language acquisition administered from 12 – 24 months (Ní Choisdealbha et al., 2023).

In the current report, we use a longitudinal approach to explore whether visual rhythmic speech is processed by the infant brain. As noted, we adapted a rhythmic syllable repetition EEG paradigm based on a 2Hz repetition rate, used previously with children in tests of TS theory (Power et al., 2012, 2013; Keshavarzi et al. 2022a). This repetition rate corresponds to

the rate of the production of stressed syllables across languages (Dauer, 1983), and is within the delta band of EEG. This is important in the context of TS theory, given that the phase alignment of delta band rhythms to syllables differs in dyslexia, as does neural encoding of delta-band speech envelope information (Power et al., 2012, 2013, 2016). The stressed syllable rate has also been shown to be strongly represented in infant EEG responses to IDS (Attaheri et al., 2022a; Menn et al., 2023). The Cambridge UK BabyRhythm study used one low-frequency repetition rate (2Hz) in multiple modalities (speech, non-speech, auditory-only, visual-only, AV). This was designed to enable estimation of cortical tracking of auditory-only, visual-only and AV speech (Power et al., 2012; Ní Choisdealbha et al., 2022, 2023). In Power et al.'s original (2012) study, typically-developing 13-year-old children showed significant tracking of visual-only speech, and the phase coherence of cortical tracking across trials was related to performance on a language task. These results contrast with those of Tan and colleagues (2022), who did not find significant visual-only speech tracking for IDS at any age in their study, which included four-year-olds and adults in addition to five-month-old infants.

Thus, simple rhythmic speech may provide an easier stimulus to align to than IDS, and consequently may enable us to examine whether individual differences in phase alignment might affect language outcomes. Further, Power et al. (2012) also reported that visually-presented speech altered the preferred phase of *auditory* entrainment. This was examined by subtracting visual-only EEG from AV responses (AV-V) in the syllable repetition paradigm and comparing the resulting values to responses to auditory-only speech. The comparison between AV-V and auditory-only speech conditions revealed a change in preferred phase at 4 Hz. Accordingly, comparing infant visual-only and AV responses in the rhythmic speech paradigm may reveal information about changes in preferred phase in infants, changes that could be relevant to the visual system preparing the auditory oscillators to be at the "optimal" phase angle for speech processing (Schroeder, 2008).

The core research question explored here was whether individual differences in phase-alignment to rhythmic visual-only speech exert an influence on later language development. A number of research questions were of interest contributing to this primary question. First, on the basis of our prior infant and child data, and the adult literature regarding oscillatory mechanisms in speech processing, we expected to find significant cortical tracking of visual-only rhythmic speech by infants. To investigate cortical tracking, we examined whether cortical power at the stimulus frequency (2Hz) increased in response to the stimuli. We also measured inter-trial coherence, which tells us whether the temporal alignment of neural oscillations at 2Hz was similar, relative to stimulus onset, across trials. Second, we were interested in the presence of preferred phase both within and across infants. For group level preferred phase, we investigated whether individuals' mean angle of alignment to the rhythmic stimuli coalesce around a similar point at a group level. For individual level preferred phase, we investigated whether individuals' mean phase angles differed when watching visual speech compared to a resting condition with no visual stimulation. These measures enabled us to investigate potential developmental changes in visual-only speech tracking and phase alignment between five and eight months of age. Third, following Power et al. (2012), we investigated whether the availability of visual speech information affects auditory cortical tracking. We did this by subtracting visual-only responses from AV responses and examining stimulus-driven cortical power and phase alignment. Following Power et al.'s results, we might expect to find a common preferred phase across infants at 4 Hz. Finally, we were interested in whether individual differences in infant phase angles (a) to visual-only speech and (b) between AV and visual-only speech would predict infant-led and parent-reported measures of language development taken at 12, 18 and 24 months of age. If the phase alignment mechanisms for processing visual speech information are intrinsic to speech processing and support language development (rather than being dependent on language comprehension) we would expect to

see cortical tracking of visual-only rhythmic speech in pre-verbal infants, as well as phase alignment of visual-only neural responses to the stimuli. We would also expect to see relations between phase alignment and later language performance.

2. Methods

2.1 Participants

Participants were infants in the longitudinal Cambridge UK BabyRhythm Project. In total, 122 infants (65 male, 57 female) were recruited to the project and 113 participated through the entire EEG data collection period, visiting the lab for assessments with the rhythmic speech paradigm at two, five, six, eight and nine months. The five- and eight-month appointments measured infant tracking of visual-only speech, at six and nine months an AV paradigm was employed. For the visual speech assessments, 36 infants met the inclusion criteria of having at least 5 trials in which they were detected by the eye-tracker as looking to the screen for at least 75% of the trial (19 infants present for both age groups). A strict looking time criterion was necessary because when a child looks away from a visual-only stimulus, this interrupts entrainment or other forms of oscillatory tracking because they are no longer receiving input to track. For the AV-V assessments, there were 37 infants (six months) and 40 infants (nine months) who watched at least five AV syllable trials for at least 75% of the trial data. However, for these AV-V analyses, only infants who also had valid data for the preceding month's visual-only paradigm appointment could be included. This left 18 infants with data at five and six months, and 22 infants with data at eight and nine months. Full reasons for exclusions are given in Appendix A, Table A1. Sample sizes for each analysis are given in Table A2 and Table A3. Data collection occurred in the United Kingdom where data on race/ethnicity are protected and cannot be collected without special cause.

An unknown technical factor affected mean phase angles that were computed for datasets collected at appointments before or after February 14th 2018 (see Ní Choisdealbha et al., 2023, for details). Due to the eye-tracking criteria, few infants tested before 14th February 2018 had valid data at five or eight months, (a minimum of two at eight months to a maximum of eight at five months), so these infants were excluded. The final number of infants included in the visual-only phase angle analysis was therefore 28 (five months), and 33 (eight months). For the AV-V analyses, there were 15 infants with data at five and six months, and 21 at eight and nine months.

2.2 Stimuli and Apparatus

Stimuli were a video of a female speaker, recorded from the shoulders up, looking at the camera and repeating the syllable "Ta" twice per second for 12 seconds. Videos were presented on a TX300 eye-tracker screen (Tobii AB, Stockholm, Sweden) which recorded eye-tracking data at a rate of 300 samples per second. At five and eight months, the stimuli were silent; at six and nine months, the repeated syllable was played through 2020i speakers (Q Acoustics Ltd., Bishop's Stortford, UK) driven by a Topaz AM5 stereo amplifier (Cambridge Audio Ltd., London, UK). The experiment was run using PsychToolbox (Brainard, 1997; Pelli, 1997; Kleiner, Brainard & Pelli, 2007) via Matlab r2018b.

EEG data were recorded using an infant-friendly 64-channel Geodesic Sensor Net (Electrical Geodesics, Inc., Eugene, OR, USA) connected to a GES 300 amplifier recording via Netstation acquisition software at a sampling rate of 1kHz. A Cedrus StimTracker was used to insert events into the EEG data to mark the onset of each "beat" of the syllable.

2.3 Procedure

Recording took place in an EEG cubicle with infants sitting in an infant chair facing the eye-tracker screen. With the EEG cap plugged in and recording, infants' eye positions were detected using the Tobii infant-friendly five-point calibration procedure. At five and eight

months, infants watched the stimulus videos consecutively. There was no defined inter-stimulus interval but a brief buffering period of variable duration between each video, in which no face was visible, meant that stimulation was not continuous. If the infant was no longer engaged by the experimental stimuli on the screen, one or more attention-grabber stimuli of varying lengths were played to re-engage their attention via novelty. At six and nine months the syllable stimuli were also interspersed with other empirical stimuli showing a ball bouncing on a drum and creating a 2Hz drumbeat sound (Ní Choisdealbha et al., 2022, 2023). Overall, infants were shown 45.5 ± 11.9 (mean ± SD) syllable videos at five months, 42.1 ± 9.7 at six months, 50.7 ± 9.9 at eight months and 45.4 ± 6.3 at nine months. Once infants had seen approximately 50 videos of the syllable or had become fussy or inattentive to the stimuli, EEG recording took place in silence with no rhythmic visual or auditory stimulation for three to five minutes, referred to hereafter as the "resting state". During this part of the recording, a researcher sat beside the infant and silently blew bubbles or pointed to images in a picture book in a non-rhythmic way. This resting state condition was consistent across all recordings in the Cambridge UK BabyRhythm project (see also Attaheri et al., 2022a; Ní Choisdealbha et al., 2023).

The decision to present the AV and visual-only stimuli at different appointments was made to limit attrition due to infant fussiness or tiredness, and maximise the quantity and quality of data collected for each condition. While there may be age-related differences in tracking and phase alignment, no age-related differences were found regarding infants' responses to AV stimuli between 6 and 9 months (Ní Choisdealbha et al., 2023). Furthermore, in the same sample, no age-related differences in stimulus tracking using the mTRF approach were found between 7 and 11 months, though there were differences between 4 and 11 months (Attaheri et al., 2022a,b). Given that our prior data indicates that cortical tracking and phase alignment is similar between 6 and 9 months, we may assume that the month-gap necessary

here in recording infant responses to visual-only and AV speech did not exert strong effects on the subtraction analyses.

2.4 EEG processing

EEG data were preprocessed and analysed as described in Ní Choisdealbha and colleagues (2022, 2023), using EEGlab 14 (Delorme & Makeig, 2004). Filtering, noise reduction (ASR; Kothe & Makeig, 2013), bad channel detection and interpolation, epoching, baseline correction and re-referencing were all performed as described in the previous manuscripts. Key details are repeated in Appendix A.

Eye-tracking data were analysed using custom scripts, to determine for what proportion of the trial the infant was looking at the screen. To account for brief losses in tracking, a conservative interpolation procedure was implemented to fill in gaps in tracking of less than 50ms during which gaze did not move more than 10% of the width or height of the screen. If the infant watched the screen for at least 75% of the length of the trial, that trial was included in further analyses (see Table 1 for mean looking times, numbers of valid trials). The resting state data was processed in the same manner as the syllable condition data, and epoched into segments of the same length (see Table 1).

2.5 EEG analysis

2.5.1 General approach

The analysis approach mirrored that of other studies from the Cambridge UK BabyRhythm project (Ní Choisdealbha et al., 2022; 2023). Starting with trial-level data, we (1) use a normalised FFT to check if there is an increase in neural power in response to the stimulus at the stimulus frequency, to establish whether tracking is taking place. We then (2) use an inter-trial coherence (ITC) measure to see if changes in power are time-locked to the stimulus. This tells us if the infant brain tracks the visual input precisely, or if the response is variable. Regarding the individual “beats” (syllables or similarly-spaced instances within the resting

state segments), we then use the time-frequency data from step two to examine (3) the mean angle of the 2Hz response recorded at the same time as each beat of the stimulus was played (phase angle). This time-frequency data is also used to explore (4), the consistency at which each "beat" of the response occurred at that angle (indexed by vector length). The angle of the 2Hz response to the visual speech stimulus in step (3) indicates whether the timing of that response lines up differently in response to the stimulus, versus during resting state. The vector length measure in step (4), like ITC, indexes consistency, but across beats rather than trials.

The final step (5) of EEG analysis moves from individual-level data to group-level data. The individual analyses in steps three and four tell us whether, within infants, there are differences in angle and consistency across conditions. The group-level preferred phase analysis (step 5) tells us if there is a common preferred phase alignment between stimulus and response, which in turn might suggest an “optimal” alignment – something which is important for the language analyses. For each of these different EEG measures (relative power [1], ITC [2], individual phase angle [3], individual vector length [4], and group phase angle [5]) we examine both responses to the visual-only stimuli, and the values obtained when the response to the visual-only stimulus is subtracted from the response to the AV stimulus.

Like other neural analyses within the Cambridge UK BabyRhythm project, we take a whole-head approach to relative power and to ITC (Attaheri et al., 2022; Ní Choisdealbha et al., 2022). The assumption is that, if tracking is taking place, the phenomenon should be robust enough to pick up even when averaging across the scalp. For the results related to phase angle ([3], [4] and [5] in the above list), it was possible that differences in phase profile across different scalp regions of interest (ROIs) would attenuate locally-relevant results (for depictions of phase angle differing by location, see Doelling et al., 2019; Ní Choisdealbha et al., 2023). Consequently, we followed previous work (Cabral-Calderin & Henry, 2022; Ní Choisdealbha et al., 2023) in using a combination of electrode-based testing and visual

inspection to identify electrodes with a stronger 2Hz response for the visual-only condition, and a stronger 2Hz response for the AV – V data, and computing phase angle over the relevant clusters. These included a mid-right frontal (MRF) and parietal (P) cluster for the visual-only analyses, and right temporo-parietal (RTP), mid-right frontal (MRF), mid-left frontal (MLF), and left temporal (LT) clusters for the AV – V analyses. Details including specific electrodes can be found in Appendix B in the supplementary materials. Topographic plots showing where the scalp-level 2Hz responses differ between visual stimulation and resting state, and between audiovisual and visual stimulation, can be found in Figure B1 in the supplementary materials.

2.5.2 Relative power

For step (1) above, data were analysed using the same approaches as used by Ní Choisdealbha et al. (2022; 2023). The relative power analysis indicates whether an increase in 2Hz power took place. A Fast Fourier Transform (FFT) was run on the data using the MTMFFT approach in Fieldtrip (Oostenveld et al., 2014) with a Hanning taper, using the 10.5-second length of the trial and with frequency bins in 0.1Hz steps between 0.5 and 15Hz. Typically, power is greater in the lower frequencies of the neural signal. To correct for this, data were then normalised to neighbouring frequency bins by subtracting the mean power of the four nearest frequency bins (two either side) from each frequency bin (Nozaradan et al., 2011). This means that results show which frequency bins had greater (or lesser) power relative to its neighbouring bins. Datapoints more than 3.5 times the standard deviation away from the mean were excluded as outliers.

To obtain the difference values for the AV-V analyses, the normalised FFT values at each frequency bin and in each condition (syllable and resting state) for the visual-only paradigm were subtracted from the corresponding values for the AV paradigm at the next age-point (i.e. five month values subtracted from six months, eight month values from nine months).

399

400 2.5.3 Inter-trial coherence (ITC)

401 For step (2) above, to find whether changes in power were time-locked to the stimulus, a time-
402 frequency analysis was run in Fieldtrip, using the wavelet method with five Morlet wavelets
403 and taking 0.1Hz steps between 1 and 15Hz. The ITC analyses show whether the timing (or
404 phase) of the 2Hz response was consistent across trials. This would indicate that the increase
405 in 2Hz power, if present, was tagged to a particular event (i.e. the onset of the stimulus). The
406 ITC approach reflects both overall increases in power as well as phase consistency.
407 Consequently, the effects of increases in power, regardless of phase, were corrected for by
408 subtracting jittered from non-jittered ITC data so that the final ITC score would reflect *phase*
409 *consistency* across trials regardless of power (this analysis was also performed in Ní
410 Choisdealbha et al., 2022, and more information is available in Appendix A of the
411 Supplementary Materials). Datapoints more than 3.5 times the standard deviation away from
412 the mean were excluded as outliers. To obtain the difference values for the AV-V analyses, the
413 same subtraction approach as for the FFT was applied to the (non-jittered minus jittered) ITC
414 values.

415 2.5.4 Phase angle and vector length

416 The values for use in analysis steps (3), (4) and (5) were computed using the Circular Statistics
417 Toolbox (Berens, 2009). The approach used for the trial-level ITC data was repeated for each
418 of the ROIs individually, and the *circ_mean* function was used to compute the mean phase
419 angle and resultant vector lengths of infants' neural responses at each analysis frequency. Phase
420 angles were observed every 500ms from 1.5 to nine seconds after trial onset. Negative mean
421 phase angle values had two times pi added to them so that all results fell between 0 and 360
422 degrees.

For the phase angle and vector length differences, the mean phase data for each infant at each age for each frequency bin for the visual-only speech was subtracted from the matched responses for the AV speech (i.e. five month values subtracted from six months, and eight months from nine months). The mean phase angle and vector length of this "phase difference" for each infant was then obtained using the *circ_mean* function. In each case (visual-only or AV-V) these values could then be compared by age, condition, ROI, and frequency bin to examine their effects on within-individual preferred phase (steps 3 and 4) or compared to a uniform distribution to examine group preferred phase (step 5; see sections 2.6.2. and 2.6.3 below).

2.6 Statistical analysis

2.6.1 Relative power and inter-trial coherence

Using lmerTest (Kuznetsova, Brockhoff & Christensen, 2017) in R, linear mixed effects models were run on the normalised FFT and ITC values. There were fixed factors of age (five versus eight months), condition (visual only versus resting state), frequency bin, and their interactions. There was a random intercept on participant identity and where it was possible to include a random slope on age without creating convergence or boundary issues, this is noted in the results section. Satterthwaite-corrected ANOVAs were run on the models to determine variables and interactions that contributed significantly to the models, while beta estimates are reported to illustrate the size and direction of effects.

In all cases, the key interaction is that between frequency bin and condition, where an increase in relative power and intertrial coherence at the 2Hz frequency bin in the syllable condition would indicate that cortical tracking occurred at the stimulus frequency. The 2 Hz, 3 Hz, 4 Hz, 5 Hz, 6 Hz and 7 Hz frequency bins were included in the models, with 1 Hz also included in the FFT model but not the ITC model (due the wavelet approach omitting values on the edges of the time-frequency range). The inclusion of multiple frequency bins for

comparison purposes allows us to examine whether the increases in power and ITC occurred at the 2Hz frequency bin specifically. While effects might be expected at the harmonic frequencies (4Hz, 6Hz), an increase in relative power or ITC would not be expected at the other frequencies. For comparability with related work (e.g. Attaheri et al., 2022a,b, Power et al., 2013) we define 1Hz, 2Hz and 3Hz as delta, and 4Hz to 7Hz as theta.

For the relative power and ITC AV-V analyses, the same analyses were run on the AV-V values. In this case the age factors were 5/6 months and 8/9 months; the conditions were the AV-V difference and the difference in corresponding resting state values between the paired recordings.

2.6.2 Vector length and phase angle

The analyses of individual-level phase angle and vector length used mixed effect regressions with factors of age, condition, frequency bin and ROI, along with their interactions. Vector length indexes whether an infant's neural responses at a given frequency occurred at the same phase angle across trials, or if and by how much they varied. Analysis of individuals' mean phase angle across conditions and frequency bins tell us whether the angle of phase alignment to the stimulus differed from the mean phase angle observed at the sampled timepoints during resting state. Phase angle is not a linear measure. High values are adjacent to low values (0° is also 360°). To account for the circular nature of the phase angle data, models were run on the sine and cosine values of each angle rather than the raw radian values. Sine and cosine provide a linear representation of the circular features of the data (see also Ní Choisdealbha et al., 2023).

2.6.3 Group-level phase angle

The individual phase angle analyses tell us whether phase angle (or rather, its sine and cosine) differed between conditions and frequency bins. The group-level analysis, conversely, tells us whether responses converged across individuals, that is, whether infants showed a similar preferred phase angle in response to the stimuli. The Rayleigh statistic was used to examine

whether mean phase angles on a *group* level converged toward a particular angle. It indicates whether a set of circular values differ from a uniform distribution, that is, whether they are clustered in a particular direction. The Rayleigh statistic for each group-level distribution of phase angles were obtained using the *circ_rstat* function in the Circular Statistics Toolbox (Berens, 2009). Based on prior work (Ní Choisdealbha et al., 2023), significant clustering of responses was expected at 2Hz for the six- and nine-month responses to the AV syllable and the same clustering was examined for the visual-only five- and eight-month responses. Following Power and colleagues (2013), 4Hz clustering was examined for the AV-V analyses.

2.7 Language development analyses

The language measures were decided *a priori* (Rocha et al., 2022 preprint, for rationale) and were grouped into parent-reported standardised measures and infant-led experimental measures encompassing phonology, communicative gesture, and vocabulary. The same measures were used for all Cambridge UK BabyRhythm papers using neural predictors to estimate language outcomes (Attaheri et al., 2022 preprint; Ní Choisdealbha et al., 2023). The parent-reported measures were receptive and productive vocabulary scores in the Lincoln CDI (Meints, Fletcher & Just, 2017) at 24 months. The infant-led measures were scores from three tasks completed by the infants: (a) a binary measure of whether the infant could point communicatively or not at 12 months; (b) their scores on the Computerised Comprehension Task (CCT; Friend & Keplinger, 2003) at 18 months, in which infants point to pictures labelled by the experimenter; and (c) their scores on a non-word repetition (NWR) task conducted at 24 months, encompassing the proportion of correctly-repeated consonants as well as accuracy in reproducing the number of syllables and the stress pattern when repeating an item. Further details on these tasks can be found in Rocha and colleagues (2022 preprint).

Relations between language outcome measures and the measures of phase alignment were assessed using circular-linear correlations between phase angle and the outcome measures

(*circ_corrcl* in the Circular Statistics Toolbox, Berens, 2009), and linear-linear correlations between vector length and the outcome measures. This follows the approaches of Power and colleagues (2013) and Keshavarzi and colleagues (2022a) in their analyses of how preferred phase relates to phonology in children with dyslexia. For all of the linear-linear correlations, Pearson correlations were used, with the exception of the pointing measure, for which point biserial correlations were run (Nagel, 2006). The alpha threshold used was adjusted based on the number of measures in each group, that is, $\alpha = 0.01$ for the five infant-led measures, and $\alpha = 0.025$ for the two parent-estimated ones. To limit the number of tests performed, one ROI was selected for analysis for each group of tests (visual-only and AV-V).

This approach diverges from our original plan for the visual-only speech data, which was to replicate the analysis strategy used in our prior neural predictor papers (Attaheri et al., preprint; Ní Choisdealbha et al., 2023). In these papers, multivariate models were run using the infant-led and parent-estimated measures, respectively, as dependent variables in each model. The multivariate approach was also planned for the current data, with separate multivariate models to be run on data from the five-month and eight-month age groups with predictor factors of phase angle, vector length and the interaction between them. However, substantial heteroscedasticity was observed when plotting the residuals of each model. Therefore, the correlation-based approach described above was adopted. The AV-V sample sizes did not provide sufficient degrees of freedom to run the multivariate models. Consequently, these data were also analysed for each language measure individually using correlations.

3. Results

As there are many different analyses, we begin with a brief overview of the key results. First, regarding *relative power* in response to *visual-only* speech, we do find greater relative power at 2Hz in response to the visual-only stimulus, relative to the resting state and to all other

frequency bins. This indicates the presence of cortical tracking, and did not interact with age. There were also differences in phase angle in response to the syllable relative to resting state at 2Hz, both in terms of phase consistency (vector length) and the sine value of the angle (predominantly over mid-right frontal electrodes).

Turning to the group-level phase analyses, the 2Hz mean phase angle does converge for the 8-month-old infants in response to the visual-only syllable stimuli, but not for the 5-month-old infants. The AV-V phase angles in response to the syllable at 8/9 months also converge, at 4Hz. Accordingly, group preferred phases are present by 8 months.

For the other AV-V analyses, we find that power at 2Hz remains above the comparison power value (taken from the resting state at the visual-only session subtracted from that recorded at the AV session). The effect at 2Hz differs from the unrelated frequency bins, but is not significantly different from that found at the harmonic bins (1Hz, 4Hz and 6Hz). There was no effect of age, and there were no effects on ITC. The crucial condition by frequency interaction was not seen for the phase angle sine value, or vector length results, and the cosine value showed only a trend ($p = 0.057$) towards a condition-by-frequency interaction. The interaction arose because individuals' mean phase angles at 2Hz differed between the syllable condition (AV-V value) and the resting state.

Finally, there were some significant relations between individual phase angle and vector length measures, and subsequent scores on the language measures. The phase angle of the response to the visual-only syllable at 8 months was related to parent-reported receptive and productive vocabulary at 24 months.

3.1 Neural responses to visual-only speech

3.1.1 Relative power

As noted in section 2.6.1, factors in the statistical analysis of relative power were condition (visual only or resting state), frequency bin, age, and their interactions. There was a random

intercept on participant identity. We first examined the beat-corrected Fast Fourier Transform. Our goal was to see if there was significantly greater relative power at 2Hz in response to the visual stimuli, relative to the resting state EEG (Figure 1, a and b), and if this effect differed by age. Table 2 provides the results of Satterthwaite-corrected tests of the fixed effects in each model for the visual-only data (i.e. did each factor or interaction contribute significantly to each model). Analysing effects of condition, frequency bin, age, and the interactions between them, we found significant contributions of both frequency bin ($F(6, 867.88) = 4.12$, $p = 0.0004$) and the frequency bin by condition interaction ($F(6, 867.88) = 3.32$, $p = 0.003$) to the model using Satterthwaite-corrected ANOVAs. The estimates show that this was driven by a greater relative power in the 2Hz frequency bin in response to the syllable, relative to the control condition (2Hz relative power significantly higher than 7Hz base case, $\beta = 0.34$, SE = 0.11, $p = 0.002$; 2Hz relative power significantly higher than all other frequency bins 1 to 6Hz, all βs = 0.24 to 0.39 , all ps < 0.05). Accordingly, significant cortical tracking was present. It did not differ by age, as there was no signification contribution of age ($F(1, 360.3) = 0.015$, $p = 0.902$), and the interactions between age and the other variables did not contribute to model fit (age by condition, $F(1, 903.4) = 1.902$, $p = 0.168$; age by frequency bin, $F(6, 867.88) = 0.326$, $p = 0.923$; age by condition by frequency bin, $F(6, 867.88) = 0.514$, $p = 0.798$). As these model contribution results suggest, there were no significant beta estimates relating to age nor to its interactions. For this and all other analyses, full model results are available in Appendix B (Table B1, Table B2, Table B3, and Table B4).

3.1.2 Inter-trial coherence

The inter-trial coherence analysis to explore whether the stimulus drew rhythmic cortical responses into a consistent phase across trials revealed no significant effects of condition, frequency bin or age (this model allowed for a random slope on age as well as the random intercept on participant identity), nor of their interactions. Thus, although the FFT results

showed an increased cortical response to the stimuli, at the stimulus frequency (i.e. cortical tracking), the ITC results suggest that the temporal consistency of this tracking varied across trials.

3.1.3 Vector length and phase angle

The same factors – age, condition, frequency bin, and their interactions – were used in the linear regressions on vector length and on phase angle (in the form of its sine and cosine values), with the inclusion of ROI as well. Moving from trial-over-trial consistency to beat-by-beat consistency, the vector lengths of infants' mean phase angles in response to each "beat" or instance of mouth opening did reveal a significant contribution of frequency bin ($F(5, 1524) = 84.63$, $p < 0.0001$), which was reflected in longer 2Hz ($\beta = 0.069$, SE = 0.031, $p = 0.024$) and shorter 3Hz ($\beta = -0.096$, SE = 0.031, $p = 0.002$) vectors relative to the 7Hz base case. There was also a significant contribution of condition ($F(1, 1569) = 13.5$, $p = 0.0002$) and a significant condition by frequency bin by ROI interaction ($F(1, 1569) = 2.565$, $p = 0.025$). The interaction arose from a significantly longer 2Hz vector in response to the syllable for the MRF base case ($\beta = 0.1$, SE = 0.042, $p = 0.017$), with a non-significant drop in consistency for the 2Hz response to the syllable over the parietal ROI ($\beta = -0.098$, SE = 0.059, $p = 0.099$). Accordingly, on an individual-by-individual basis, condition (visual-only versus resting state) did affect the consistency of the 2Hz neural response.

Satterthwaite-corrected ANOVAs on phase angle (sine and cosine values) revealed a significant effect of condition on the sine values ($F(1, 1377) = 6.67$, $p = 0.01$). However, the only significant beta estimate was a three-way interaction between condition, ROI, and frequency bin ($\beta = 0.696$, SE = 0.35, $p = 0.047$) – suggesting that the size of the difference in the angle of individuals' 2Hz phase response to the syllable (versus during resting state) differs by scalp region. The cosine model had significant contributions of the condition by ROI, frequency bin by ROI, and condition by age by ROI interactions (Table 2). However, the

598 critical frequency bin by condition interactions were all non-significant, as were the beta
599 estimates. Nonetheless, the sine effect shows that observed phase angle at 2Hz differed between
600 the resting state and syllable conditions.

601 ***3.2 Group-level phase angle***

602 Group-level Rayleigh analyses of clustering of the phase angles suggested that the visual-only
603 speech drew responses into a similar phase angle across infants at 8 months. For resting state
604 data, we would not expect any clustering of the 2Hz phase angles given the absence of a
605 stimulus, and indeed no significant clustering was found (5 months, n = 28: MRF, Z = 1.663,
606 p = 0.191, P, Z = 1.928, p = 0.146; 8 months, n = 33: MRF, Z = 0.333, p = 0.72, P, Z = 0.302,
607 p = 0.743). By contrast, there was significant group-level clustering for the syllable at eight
608 months in the parietal ROI (Z = 3.206, p = 0.039, n = 33; Figure 2), although not in the MRF
609 ROI (Z = 0.201, p = 0.82) nor at five months (MRF: Z = 0.0.578, p = 0.565, P: Z = 1.565, p =
610 0.21, n = 28). These group-level results suggest that, by 8 months, viewing visual-only speech
611 is moving oscillatory responses towards a common preferred phase angle over parietal
612 electrodes. Although the non-significant Rayleigh tests suggest that individuals' preferred
613 phase differs over the MRF ROI, the vector length result for the syllable at 2Hz over this ROI
614 indicates that there is nonetheless a preferred phase within individuals.

615 For the AV-V data, the 4Hz band was analysed in accordance with the findings of
616 Power and colleagues (2013). The assumption is that the AV-V values represent the response
617 to an AV stimulus corrected for the response to the visual stimulus; thus the AV-V values
618 reflects the response to auditory stimulation alongside any effects related to audio-visual
619 integration. The only ROI showing group-level clustering was the MRF ROI, for the 8/9-month
620 values in the 4Hz syllable condition (Z = 4.207, p = 0.013; other 2Hz and 4Hz values reported
621 in Appendix B, Table B5). This suggests that for the 8/9-month-olds, auditory responses in the

presence of visual stimulation are drawn towards a common 4 Hz phase. This indicates that there is *auditory* phase alignment in the theta band while watching an audiovisual stimulus.

3.3 Difference between AV and visual-only speech

3.3.1 Relative power

For the AV-V FFT model (allowing for a random slope on age group), Satterthwaite-corrected ANOVAs on the model revealed significant factors of condition ($F(1, 467.68) = 3.9$, $p = 0.049$) and a condition by frequency bin interaction ($F(5, 668.67) = 2.28$, $p = 0.045$; see Table 3 for all factors and all models). Estimates reveal a significant increase in power in response to the syllable (relative to resting state) for the 2Hz base case ($\beta = 0.4$, SE = 0.15, $p = 0.009$). The 2Hz syllable power was greater than at 3, 5 and 7 Hz (βs = -0.65, -0.44, -0.527, p's = 0.003, 0.043, and 0.016 respectively) but not than 1Hz ($\beta = -0.26$, $p = 0.24$), 4Hz ($\beta = -0.2$, $p = 0.37$) or 6Hz ($\beta = -0.36$, $p = 0.095$; Figure 1, c and d). There were no effects of age group. Figure 3 depicts the 2Hz relative power values for each of the five-, six-, eight- and nine-month sessions, showing the difference between AV and visual responses to the speech stimulus. All simple effects are reported in Appendix B, Table B4. Although the 2Hz peak in Figure 1(c), corresponding the the 5/6-month AV-V, appears larger than that in Figure 1(d) (8/9-month AV-V), Figure 3 highlights why this age-related difference is non-significant in the model – a couple of infants showed very high relative 2Hz power at the younger ages (see Figure 3 caption for note on outliers).

The ITC (allowing for random slope on age group), phase angle (sine and cosine) and vector length models were then run. The vector length, sine and cosine models all had significantly contributing factors according to the Satterthwaite-corrected ANOVA (see Table 3). These significant factors were the frequency bin for vector length ($F(5, 1525) = 4.141$, $p = 0.001$), the condition by age interaction for sine values ($F(1, 1545) = 6.512$, $p = 0.011$), and a marginal effect of condition for cosine values ($F(1, 1543) = 3.779$, $p = 0.052$ (see Table 3 for

all values). The crucial condition by frequency interaction did not make a significant contribution to any model. Simple effects showed only a marginal interaction between the 2Hz frequency and the syllable condition for the left temporal base case in the cosine model (β = 0.688, SE = 0.361 p = 0.057).

3.4 Language development analyses

The results of the circular-linear and linear (Pearson or biserial) correlations with the visual-only phase angle data are reported in Table 4. The MRF ROI was used in these analyses as this is where we saw greater phase consistency within individuals. We note that there is a well-documented right hemispheric preference for encoding rhythm (Sammler et al., 2015) especially at slower rates (Vanvooren et al., 2014). Significant correlations were found between phase angle in response to the stimulus at 8 months, and both receptive ($\rho(26) = 0.536$, p = 0.018) and productive vocabulary ($\rho(26) = 0.547$, $p = 0.015$) at 24 months, suggesting that preferred phase may play a specific role in vocabulary development. The correlations with vector length were all non-significant, suggesting that it may be the angle of phase alignment, and not its consistency, which plays a role in the relation between visual speech processing and later language acquisition.

The AV-V results relating to ROI were uninformative, hence we used the LT ROI for the analyses regarding subsequent language development given that this region was used for language analyses in related work (Ní Choisdealbha et al., 2023). Circular-linear and linear (Pearson and biserial) correlations were run on the 5/6-month-olds' data and the 8/9-month-olds' data. In these analyses, there were no significant associations for the infant-led measures nor the parent-reported CDI scores. The closest trends related to the 5/6-month-olds' AV-V vector lengths, and the 24-month NWR syllable-matching (p = 0.038) and stress-matching (p = 0.031) measures, also the 8/9-month phase angle and 18-month CCT scores ($\rho(16) = 0.61$, p

= 0.037). Results for all measures are presented in Table 5 and should be interpreted in light of the small sample sizes.

Discussion

Here we explored whether 'visual prosody' (low-frequency visual temporal information in speech) evokes cortical tracking in five- and eight-month-old infants, whether individual differences in preferred phase to visual-only speech affect subsequent language development, and whether group preferred phase angles change when visual-only speech EEG is subtracted from AV speech EEG (AV-V). The rhythmic speech paradigm employed here revealed that infants show significant cortical tracking of visual-only speech at both five and eight months, and that individuals do show phase consistency (longer phase angle vectors) in response to the visual syllable. Convergence towards a group *preferred phase angle* for visual-only speech was not significant at five months, but a preferred phase angle for visual-only speech had emerged by eight months. Analyses of the difference in neural response between visual-only and AV speech at 5/6 and 8/9 months revealed convergence towards a group *preferred phase angle* at 4 Hz rather than 2 Hz during cortical tracking, again for the older infants only. Consistent with the adult and animal literature, therefore, the pre-verbal infant brain shows automatic use of both visual and auditory prosodic information during speech processing. Further, individual differences in these automatic neural mechanisms have consequences for language acquisition. Each of these conclusions is discussed in greater detail below.

Regarding cortical tracking, the rhythmic speech paradigm employed here found significant cortical tracking of visual-only speech at both ages tested, shown by the normalised FFT results. The neural data from our 5-month-old infants contrasts with the negative result for cortical tracking of visual-only speech for 5-month-old infants reported by Tan et al. (2022). This could be due to some differences between the studies, such as the stimuli employed (repeated syllables versus IDS) or attention thresholds (75% versus 15%). If we had utilised a

lower threshold in the current work, this would affect the subtractions between AV and visual-only tracking, as an infant who disengages from an AV trial nonetheless continues to receive rhythmic auditory input, whereas as infant who disengages from a visual-only trial no longer receives any input. A final difference with the work by Tan et al. (2022) is that they used the mTRF approach to correlate the neural response with the envelope of the presented speech, while our focus was on individual and group-level *preferred phase*.

Regarding the phase analyses, we did not find significant ITC at either age, suggesting that while cortical tracking occurred in response to the visual-only speech, the 2Hz neural response was not drawn into a consistent alignment relative to stimulus onset for individual infants across all trials. However, when we look at phase responses to individual stimulus beats, broken down by region of interest, we see within-participant consistency in the vector lengths of the 2Hz response to the visual syllable. The analyses of the circular data showed that phase consistency was evident at the group level by 8 months. The group-level consistency manifested over posterior, parietal electrodes while the individual-level consistency was stronger over anterior, mid-right frontal electrodes. Activity at the scalp does not necessarily reflect activity over adjacent cortex. However, it may be that the posterior response reflects the evoked visual response, which is a fundamental sensory-perceptual response, while the mid-right frontal response reflects other aspects of rhythmic processing. Accordingly, there may be individual differences across infants even if consistency occurs within individuals.

As shown by these analyses, individual infants' mean 2Hz phase angles in response to visual-only speech clustered in a specific direction, suggesting that there is a *preferred phase* at which the infant brain processes 2Hz visual input, both individually and across the group. Viewing visual-only speech thus affected the *timing* of the neural response, possibly by re-setting the auditory cortex to the "optimal" state for processing anticipated succeeding vocalisations (Schroeder et al., 2008). Regarding the apparently later emergence of group-

level preferred phase effects (older infants only) compared to cortical tracking effects (five and eight months) it is notable that behavioural studies of infant speech processing using looking measures or eye tracking measures indicate a change in preference between four and eight months of age from looking at the eyes of a speaker to looking at the mouth (Lewkowicz & Hansen-Tift, 2012; Pons, Bosch & Lewkovicz, 2015). The increased focus on the mouth as infants get older is thought to provide infants with enhanced audiovisual cues during language learning (Lewkowicz & Hansen-Tift, 2012). Such cues may include the visual prosody of the speech directed to infants, who can detect such prosody from facial and head movements (Kitamura et al., 2014). Accordingly, one speculative explanation for the significant group-level preferred phase to visual-only speech found here at eight months is that it could reflect, at least in part, behavioural changes in looking preference.

Turning to the AV-V analyses, the FFT statistics revealed significant tracking of the rhythmic auditory stimulation in the presence of visual input. However, the individual-level phase results were non-significant, suggesting that the alignment of neural response to stimulus was variable. Nonetheless, the Rayleigh statistics showed that at 8/9 months (though not 5/6 months), 4 Hz AV-V angles converged towards a common direction, indicating a group-level preferred phase. This AV-V result partially replicates earlier findings with older children. Power and colleagues (2012) found 4 Hz responses to the rhythmic speech stimulus for AV and auditory-only speech, but not visual-only speech. It is notable that in the previous analysis of the infant AV-only speech data (reported by Ní Choisdealbha et al., 2022; these analyses included additional six- and nine-month-old infants who did not meet the looking time threshold used for the current paper), there was a significant increase in power at 4 Hz in response to the AV syllable. Thus, auditory stimulation at 2 Hz – in the presence or absence of visual stimulation – has effects at other, related frequencies, which visual stimulation alone does not. A simple although speculative explanation could be that the auditory component of

AV speech contains 4 Hz information that is not carried by the visual input, but that is potentially modulated by the visual input. One interpretation could be that for both infants and 13-year-olds (Power et al., 2012), 4 Hz AV-V effects are indicative of the re-setting of the auditory cortex to the "optimal" state for processing the succeeding vocalisations by visual speech information (Schroeder et al., 2008). Alternatively, the significant theta band responses could reflect the automatic coupling of the auditory system to the motor system when speech is heard, an effect indicated by adult auditory-only rhythmic speech data collected by Assaneo and Poeppel (2018). Both of these speculative explanations could be examined in future research.

Regarding subsequent language acquisition, if cortical tracking of visual prosody plays a key role in neural speech processing, then infant responses to visual-only speech should be predictive of later language outcomes. Individual differences in phase alignment to visual-only speech significantly affected subsequent language development for some measures. At eight months, individuals' preferred phase to visual-only speech predicted parental estimates of both receptive and productive vocabulary development at 24 months as measured by the UK-CDI (Meints, Fletcher & Just, 2017). In prior analyses with these same infants, phase angle in response to the AV syllable at nine months also significantly affected 24-month UK-CDI scores (Ní Choisdealbha et al., 2023). If the role of visual speech is to promote phase-resetting for optimal auditory processing, then we would expect AV-V phase responses – reflecting auditory processes in the presence of visual speech – to also predict later language acquisition. We did not find any robust relations between AV-V phase and language acquisition in this study. However, some trends were present which could be explored in a larger sample. Particularly interesting in the light of Power et al.'s (2013) data were those related to the NWR task, which is a speech production measure, and can also be considered a measure of phonological

processing skills when used with very young children (Moritz et al., 2013; Kalashnikova et al., 2021).

Our study has some limitations. There are myriad interesting questions about how the tracking and processing of visual speech develops, including identifying the cortical sources of such tracking in auditory and visual areas of the brain, and whether purported phase alignment mechanisms are specific to visual speech or apply to other forms of visual or AV stimuli. These questions were beyond the scope of the current study. A second limitation is that, despite a large longitudinal EEG sample of 113 infants, there was substantial participant attrition. This was primarily driven by the requirement for participants to have clean, usable EEG data for at least five trials in which they attended to the screen for at least 75% of a 12-second video. Furthermore, for the AV-V analyses, infants had to meet this criteria for two consecutive testing sessions. Such attrition is a common finding in infant studies (see discussion in Tan et al., 2022). Nevertheless our EEG analysis sample sizes of 18 (5/6-month AV-V) to 33 (eight-month visual only) are representative of the infant literature (e.g. there were 18 five-month-olds in Tan et al.). Given the parallels between our findings regarding preferred phase and those reported by Power et al. (2012), the results obtained with the rhythmic repetition paradigm may be treated as robust despite reduced sample sizes. A third limitation is that sample size was particularly affected in the analyses predicting language outcomes, for which EEG and language data from multiple testing sessions per infant was required. Nonetheless, many of the effects of phase angle and vector length on later language performance reported here are comparable to effects from the AV and auditory-only paradigms used in other test sessions with the same infants, which employed larger samples (>100 infants, Ní Choisdealbha et al., 2023). A final limitation is that the rhythmic speech paradigm is not comparable to connected speech, although it shares some key acoustic parameters with IDS (Leong et al., 2017). While the rhythmic speech paradigm is well-suited to phase analyses, of central interest here, it would

795 also be interesting to examine phase alignment during visual-only and AV-V speech processing
796 by infants when IDS is the input. In summary, our results show that neural oscillatory
797 mechanisms revealed by adult language processing studies, such as visual-only cortical
798 tracking (Park et al., 2016), appear to be engaged during visual speech processing in the first
799 year of life. These neural mechanisms are also engaged during auditory processing in the
800 presence of visual speech, with cross-modal modulation effects in evidence, and predict
801 language outcomes.
802
803 *Conclusions*
804 Adult studies of speech processing have suggested that auditory and visual low-frequency
805 neuronal oscillations act in concert when human speech is the input, with cortical tracking of
806 both auditory and visual prosody enhancing speech comprehension. The rhythmic speech
807 paradigm used here with infants reveals some similar visual and auditory oscillatory
808 mechanisms in the first year of life, when the infant is still pre-verbal. Whether the
809 underlying neural mechanisms measured here are general sensory-perceptual mechanisms or
810 specific to speech and language is not a question that we address here. Instead, we show that
811 individual differences in the angle and consistency of phase responses to the rhythmic
812 syllable stimuli were associated with certain measures of language development.
813 Consequently, individual differences in the engagement of these neural mechanisms by pre-
814 verbal infants, which are automatic mechanisms and not under conscious control, appears to
815 have potential implications for the early processing and development of language.
816 Accordingly, some of the neural mechanisms revealed by adult studies may be intrinsic to the
817 neural speech processing architecture, rather than dependent on learning to comprehend and
818 produce speech.

References

Assaneo, M. F., & Poeppel, D. (2018). The coupling between auditory and motor cortices is rate-restricted: Evidence for an intrinsic speech-motor rhythm. *Science Advances*, *4*(2), aao3842.

Attaheri, A., Ní Choisdealbha, Á., Di Liberto, G. M., Rocha, S., Brusini, P., Mead, N., Olawole-Scott, H., Boutris, P., Gibbon, S., Williams, I., & Goswami, U. (2022a). Delta-and theta-band cortical tracking and phase-amplitude coupling to sung speech by infants. *NeuroImage*, 247, 118698.

Attaheri, A., Panayiotou, D., Philips, A., Ní Choisdealbha, Á., Di Liberto, G.M., Rocha, S., Brusini, P., Mead, N., Flanagan, S., Olawole-Scott, H., & Goswami, U. (2022b). Cortical tracking of sung speech in adults vs infants: A Developmental Analysis. *Frontiers in Neuroscience, 16*, 842447.

Attaheri, A., Ní Choisdealbha, Á., Rocha, S., Brusini, P., Di Liberto, G. M., Mead, N., ... & Goswami, U. (2022 preprint). Infant low-frequency EEG cortical power, cortical tracking and phase-amplitude coupling predicts language a year later. *bioRxiv*, 2022-11.

Berens, P. (2009). CircStat: a MATLAB toolbox for circular statistics. *Journal of Statistical Software*, *31*, 1-21.

Bergelson, E., & Swingley, D. (2012). At 6–9 months, human infants know the meanings of many common nouns. *Proceedings of the National Academy of Sciences*, *109*(9), 3253-3258.

Bourguignon, M., Baart, M., Kapnoula, E. C., & Molinaro, N. (2020). Lip-reading enables the brain to synthesize auditory features of unknown silent speech. *Journal of Neuroscience*, *40*(5), 1053-1065.

Brainard, D. H. (1997). The Psychophysics Toolbox. *Spatial Vision, 10,* 433-436.

844 Cabral-Calderin, Y. & Henry, M.J. (2022). Reliability of neural entrainment in the human
845 auditory system. *Journal of Neuroscience, 42 (5),* 894-908.
846 Chandrasekaran, C., Trubanova, A., Stillittano, S., Caplier, A., & Ghazanfar, A.A. (2009).
847 The natural statistics of audiovisual speech. *PLoS Computational Biology, 5(7),*
848 e1000436.
849 Dauer, R. M. (1983). Stress-timing and syllable-timing reanalyzed. *Journal of Phonetics,*
850 *11(1),* 51–62.
851 Delorme, A., & Makeig, S. (2004). EEGLAB: an open source toolbox for analysis of single-
852 trial EEG dynamics including independent component analysis. *Journal of*
853 *Neuroscience Methods*, *134*(1), 9-21.
854 Doelling, K. B., Assaneo, M. F., Bevilacqua, D., Pesaran, B., & Poeppel, D. (2019). An
855 oscillator model better predicts cortical entrainment to music. *Proceedings of the*
856 *National Academy of Sciences*, *116*(20), 10113-10121.
857 Friend, M., & Keplinger, M. (2003). An infant-based assessment of early lexicon acquisition.
858 *Behavioral Research Methods*, *25*, 302-309.
859 Giraud, A. L., & Poeppel, D. (2012). Cortical oscillations and speech processing: emerging
860 computational principles and operations. *Nature neuroscience*, *15*(4), 511-517.
861 Goswami, U. (2019). Speech rhythm and language acquisition: an amplitude modulation
862 phase hierarchy perspective. *Annals of the New York Academy of Sciences*, *1453*(1),
863 67-78.
864 Goswami, U. (2011). A temporal sampling framework for developmental dyslexia. *Trends in*
865 *Cognitive Sciences*, *15*(1), 3-10.
866 Grant, K.W., & Braida, L.D. (1991). Evaluating the articulation index for auditory–visual
867 input. *Journal of the Acoustical Society of America*, *89,* 2952-2960.

Grant K.W., & Seitz, P-F. (2000). The use of visible speech cues for improving auditory detection of spoken sentences. *Journal of the Acoustical Society of America, 108,* 1197-1208.

Kalashnikova, M., Peter, V., Di Liberto, G. M., Lalor, E. C., & Burnham, D. (2018). Infant-directed speech facilitates seven-month-old infants' cortical tracking of speech. *Scientific Reports, 8*(1), 1-8.

Kalashnikova, M., Burnham, D., & Goswami, U. (2021). Rhythm discrimination and metronome tapping in 4-year-old children at risk for developmental dyslexia. *Cognitive Development, 60,* 101129.

Keshavarzi, M., Mandke, K., Macfarlane, A., Parvez, L., Gabrielczyk, F., Wilson, A., & Goswami, U. (2022a). Atypical delta-band phase consistency and atypical preferred phase in children with dyslexia during neural entrainment to rhythmic audio-visual speech. *NeuroImage: Clinical, 35,* 103054.

Keshavarzi, M., Mandke, K., Macfarlane, A., Parvez, L., Gabrielczyk, F., Wilson, A., Flanagan, S., Goswami, U. (2022b). Decoding of speech information using EEG in children with dyslexia: Less accurate low-frequency representations of speech, not "Noisy" representations. *Brain and Language, 235, 105198.*

Kitamura, C., Guellaï, B., & Kim, J. (2014). Motherese by eye and ear: Infants perceive visual prosody in point-line displays of talking heads. *PLoS One, 9*(10), e111467.

Kleiner, M., Brainard, D., & Pelli, D. (2007). What's new in Psychtoolbox-3? *Perception, 36,* ECVP Abstract Supplement.

Kothe, C. A., & Makeig, S. (2013). BCILAB: a platform for brain–computer interface development. *Journal of Neural Engineering, 10*(5), 056014.

Kuhl, P. K., & Meltzoff, A. N. (1982). The bimodal perception of speech in infancy. *Science 218,* 1138-1141.

Kuznetsova, A., Brockhoff, P.B., & Christensen, R.H.B. (2017). lmerTest Package: Tests in linear effects models. Journal of Statistical Software, *82*(13), 1-26.

Leong, V., Kalashnikova, M., Burnham, D., & Goswami, U. (2017). The temporal modulation structure of infant-directed speech. *Open Mind, 1*, 78-90.

Lewkowicz, D. J., & Hansen-Tift, A. M. (2012). Infants deploy selective attention to the mouth of a talking face when learning speech. *Proceedings of the National Academy of Sciences*, *109*(5), 1431-1436.

Luo, H., Liu, Z., & Poeppel, D. (2010). Auditory cortex tracks both auditory and visual stimulus dynamics using low-frequency neuronal phase modulation. *PLoS Biology*, *8*, e1000445.

Mehler, J., Jusczyk, P., Lambertz, G., Halsted, N., Bertoncini, J., and Amiel-Tison, C. (1988). A precursor of language acquisition in young infants. *Cognition*, 29, 143-178.

Meints, K., Fletcher, K., & Just, J. (2017). The Lincoln Toddler Communicative Development Inventory - A UK adaptation of the MacArthur-Bates Communicative Development Inventory: Words and sentences (Toddler Form). https://cpb-eu-w2.wpmucdn.com/blogs.lincoln.ac.uk/dist/b/6736/files/2017/11/Lincoln_toddler_cdiv2-2.pdf. Accessed 18/11/2022.

Menn, K. H., Michel, C., Meyer, L., Hoehl, S., & Männel, C. (2022). Natural infant-directed speech facilitates neural tracking of prosody. *NeuroImage*, *251*, 118991.

Moritz, C., Yampolsky, S., Papadelis, G., Thomson, J., & Wolf, M. (2013). Links between early rhythm skills, musical training and phonological awareness. *Reading & Writing, 26*, 739 – 769.

Munhall, K.G., Jones, J.A., Callan, D.E., Kuratate, T., & Vatikiotis-Bateson, E. (2004). Visual prosody and speech intelligibility: head movement improves auditory speech perception. *Psychological Science, 15*, 133-137.

Munhall, K.G., Kross, C., & Vatikiotis-Bateson, E. (2002). Audiovisual perception of band-pass filtered faces. *Journal of the Acoustical Society of Japan, 21,* 519-520.

Nagel, F. (2006). Point biserial correlation. *MATLAB Central File Exchange.*

Ní Choisdealbha, Á., Attaheri, A., Rocha, S., Mead, N., Olawole-Scott, H., Brusini, P., Gibbon, S., Boutris, P., Hines, D., Grey, C., Flanagan, S., & Goswami, U. (2022). Cortical oscillations in pre-verbal infants track rhythmic speech and non-speech stimuli. In Y. Gong & F. Kpogo, (Eds.), *BUCLD 46: Proceedings of the 46th annual Boston University Conference on Language Development: Volume 2* (pp. 574-585). Cascadilla Press.

Ní Choisdealbha, Á., Attaheri, A., Rocha, S., Mead, N., Olawole-Scott, H., Brusini, P., ... & Goswami, U. (2023). Neural phase angle from two months when tracking speech and non-speech rhythm linked to language performance from 12 to 24 months. *Brain and Language, 243,* 105301.

Nozaradan, S., Peretz, I., Missal, M., & Mouraux, A. (2011). Tagging the neuronal entrainment to beat and meter. *Journal of Neuroscience, 31*(28), 10234-10240.

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

Ortiz Barajas, M. C., Guevara, R., & Gervain, J. (2021). The origins and development of speech envelope tracking during the first months of life. *Developmental Cognitive Neuroscience, 48,* 100915.

Park, H., Kayser, C., Thut, G., & Gross, J. (2016). Lip movements entrain the observers' low-frequency brain oscillations to facilitate speech intelligibility. *eLife, 5,* e14521.

Pelli, D. G. (1997). The VideoToolbox software for visual psychophysics: Transforming numbers into movies, *Spatial Vision, 10,* 437-442.

943 Poeppel, D. (2003). The analysis of speech in different temporal integration windows:
944 cerebral lateralization as 'asymmetric sampling in time'. *Speech Communication.* 41,
945 245-255.
946 Pons, F., Bosch, L., & Lewkowicz, D. J. (2015). Bilingualism modulates infants' selective
947 attention to the mouth of a talking face. *Psychological Science, 26*(4), 490-498.
948 Pons, F., Lewkowicz, D.J., Soto-Faraco S., & Sebastián-Gallés, N. (2009). Narrowing of
949 intersensory speech perception in infancy. *PNAS*, 106, 10598-10602.
950 Power, A. J., Colling, L. J., Mead, N., Barnes, L., & Goswami, U. (2016). Neural encoding of
951 the speech envelope by children with developmental dyslexia. *Brain and Language*,
952 *160*, 1-10.
953 Power, A. J., Mead, N., Barnes, L., & Goswami, U. (2012). Neural entrainment to
954 rhythmically presented auditory, visual, and audio-visual speech in children. *Frontiers*
955 *in Psychology*, *3*, 216.
956 Power, A. J., Mead, N., Barnes, L., & Goswami, U. (2013). Neural entrainment to rhythmic
957 speech in children with developmental dyslexia. *Frontiers in Human Neuroscience*, *7*,
958 777.
959 Rocha, S., Ní Choisdealbha, Á., Attaheri, A., Mead, N., Olawole-Scott, H., Grey, C.,
960 Williams, I., Gibbon, S., Boutris, P., Brusini, P. & Goswami, U.C. (preprint).
961 Language acquisition in the longitudinal BabyRhythm cohort. psyarxiv.com/28c35
962 Sammler, D., Grosbras, M. H., Anwander, A., Bestelmeyer, P. E., & Belin, P. (2015). Dorsal
963 and ventral pathways for prosody. *Current Biology*, *25*(23), 3079-3085.
964 Schroeder, C. E., Lakatos, P., Kajikawa, Y., Partan, S., and Puce, A. (2008). Neuronal
965 oscillations and visual amplification of speech. *Trends in Cognitive Science, 12*, 106-
966 113.

Sumby W. H, Pollack I. 1954. Visual contribution to speech intelligibility in noise. *Journal of the Acoustical Society of America, 26,* 212-215.

Tan, S. J., Kalashnikova, M., Di Liberto, G. M., Crosse, M. J., & Burnham, D. (2022). Seeing a talking face matters: The relationship between cortical tracking of continuous auditory-visual speech and gaze behaviour in infants, children and adults. *NeuroImage, 256,* 119217.

Vanvooren, S., Poelmans, H., Hofmann, M., Ghesquiere, P., & Wouters, J. (2014). Hemispheric asymmetry in auditory processing of speech envelope modulations in prereading children. *Journal of Neuroscience, 34*(4), 1523-1529.

Walden, B.E., Prosek, R.A., Montgomery, A.A., Scherr, C.K., & Jones, C.J. (1977). Effects of training on the visual recognition of consonants. *Journal of Speech and Hearing Research, 20,* 130-145.

Yehia, H., Kuratate, T., & Vatikiotis-Bateson, E. (2002). Linking facial animation, head motion and speech acoustics. *Journal of Phonetics, 30*, 555-568.

Yehia, H., Rubin, P., & Vatikiotis-Bateson, E. (1998). Quantitative association of vocal-tract and facial behaviour. *Speech Communication, 26*, 23-43.

992

993

994 **Table 1:** Infant looking behaviour as measured by eye tracking at the different ages.

	5 months	6 months	8 months	9 months
Average % looking to each trial	38.49% (SD = 21.18%)	40.06% (SD = 15%)	38.05% (SD = 17.71%)	39.66% (SD = 14.67%)
Average trials watched for 75%+ of trial length	4.77 (SD = 6.21)	5.01 (SD = 5.12)	5.34 (SD = 6.46)	4.61 (SD = 4.91)
Average trials watched for 75%+ of trial length (babies with 5+ valid trials only)	11.53 (SD = 5.53)	10 (SD = 4.31)	12.69 (SD = 5.51)	9.5 (SD = 4.16)
Average number of resting state segments per infant who met the inclusion criteria (5+ syllable trials with 75%+ looking)	14.32 (SD = 8.56)	15.06 (SD = 6.62)	20.06 (SD = 5.62)	16.34 (SD = 6.24)

995
996

997 **Table 2:** Satterthwaite-corrected ANOVAs on model effects, neural responses to visual-only

998 speech

	FFT	ITC	PA: cosine	PA: sine	VL

Condition	F(1, 903.41) = 2.738#	F(1, 761.37) = 0.171	F(1, 1377) = 1.967	**F(1, 1377) = 6.673****	**F(1, 1569) = 13.5*****
Frequency bin	**F(6, 867.88) = 4.12*****	F(5, 734.52) = 0.332	F(5, 1377) = 0.185	F(5, 1377) = 0.296	**F(5, 1525) = 84.632*****
Age	F(1, 360.3) = 0.015	F(1, 54.53) = 0.031	F(1, 447) = 0.125	F(1, 478) = 0.117	F(1, 39) = 0.002
ROI			F(1, 1377) = 1.803	F(1, 1377) = 0.227	F(1, 1525) = 1.539
Condition * Frequency	**F(6, 867.88) = 3.318****	F(5, 734.52) = 0.362	F(5, 1377) = 0.376	F(5, 1377) = 0.241	*F(5, 1525) = 1.996#*
Condition * Age	F(1, 903.41) = 1.902	F(1, 761.41) = 0.0013	F(1, 1377) = 0.388	F(1, 1377) = 0.366	F(1, 1569) = 0.263
Condition * ROI			**F(1, 1377) = 7.425****	F(1, 1377) = 0.045	F(1, 1525) = 0.487
Frequency * Age	F(6, 867.88) = 0.326	F(5, 734.52) = 0.81	F(5, 1377) = 0.274	F(5, 1377) = 0.121	F(5, 1525) = 0.137
Frequency * ROI			**F(5, 1377) = 2.523***	F(5, 1377) = 1.346	F(5, 1525) = 0.979
Age * ROI			*F(1, 1377) = 2.724#*	F(1, 1377) = 0.058	F(1, 1525) = 0.004
Condition * Frequency * Age	F(6, 867.88) = 0.514	F(5, 734.52) = 0.025	F(5, 1377) = 1.415	F(5, 671) = 0.76	F(5, 1525) = 1.671
Condition * Frequency * ROI			F(5, 1377) = 0.681	F(5, 1377) = 0.761	**F(5, 1525) = 2.565***
Condition * Age * ROI			**F(1, 1377) = 11.162*****	*F(1, 1377) = 3.809#*	F(1, 1525) = 0.093
Frequency * Age * ROI			F(5, 1377) = 1.143	F(5, 1377) = 0.492	F(5, 1525) = 0.314
Condition * Frequency * Age * ROI			F(5, 1377) = 0.478	F(5, 1377) = 0.941	F(5, 1525) = 0.237

# p < 0.1 * p < 0.05 ** p < 0.01 *** p < 0.001
PA = phase angle; VL = vector length

Table 3: Satterthwaite-corrected ANOVAs on model effects, neural responses to audviovisual-minus-visual (AV-V) speech

	FFT	ITC	PA: cosine	PA: sine	VL
Condition	**F(1, 467.68) = 3.895***	F(1, 419.15) = 2.798#	*F(1, 1543) = 3.779#*	F(1, 1546) = 0.526	F(1, 1544) = 0.7573
Frequency bin	F(6, 446.91) = 1.579	F(5, 410.11) = 1.889#	F(5, 1522) = 0.489	F(5, 1526) = 0.441	**F(5, 1525) = 4.141****
Age	F(1, 20.75 = 0.316	F(1, 31.09) = 1.048	F(1, 15) = 0.032	F(1, 20) = 1.158	F(1, 18) = 0.039
ROI			F(3, 1522) = 0.949	F(3, 1526) = 1.027	F(3, 1525) = 1.243
Condition * Frequency	**F(6, 446.91) = 2.198***	F(5, 410.11) = 1.71	F(5, 1522) = 1.574	F(5, 1526) = 1.181	F(5, 1525) = 0.23
Condition * Age	F(1, 467.6) = 0.514	F(1, 419.44) = 0.876	F(1, 1543) = 0.934	**F(1, 1546) = 6.512***	F(1, 1544) = 0.906
Condition * ROI			F(3, 1522) = 0.776	F(3, 1526) = 1.477	F(3, 1525) = 1.603
Frequency * Age	F(6, 446.91) = 0.233	F(5, 410.09) = 0.749	F(5, 1522) = 0.418	F(5, 1526) = 0.173	F(5, 1525) = 0.156
Frequency * ROI			F(15, 1522) = 0.749	F(15, 1526) = 0.435	F(15, 1525) = 0.826
Age * ROI			F(3, 1522) = 0.921	F(3, 1526) = 0.692	F(3, 1525) = 0.493
Condition * Frequency * Age	F(6, 446.91) = 0.304	F(5, 410.09) = 0.373	F(5, 1522) = 0.764	F(5, 1526) = 0.996	F(5, 1525) = 1.021
Condition * Frequency * ROI			F(15, 1522) = 0.92	F(15, 1526) = 0.75	F(15, 1525) = 1.105
Condition * Age * ROI			F(3, 1522) = 0.276	F(3, 1526) = 1.795	F(3, 1525) = 0.183
Frequency * Age * ROI			F(15, 1522) = 0.526	F(5, 1526) = 0.91	F(15, 1525) = 1.423
Condition * Frequency * Age * ROI			F(15, 1522) = 0.992	F(5, 1526) = 0.474	F(15, 1525) = 0.909

# p < 0.1 * p < 0.05 ** p < 0.01 *** p < 0.001
PA = phase angle; VL = vector length

Table 4: All circular-linear (phase angle) and linear (vector length) correlations between visual only phase responses over mid-right frontal ROI, and language measures. Alpha is corrected

1009 for multiple comparisons and is equal to 0.01 for the infant-led and 0.025 for the parent-
1010 estimated measures, accounting for five different infant-led measures and two different parent-
1011 reported vocabulary measures.

	Circular-linear correlations with phase angle		Linear correlations with vector length	
	5 months	8 months	5 months	8 months
Infant-led measures				
Pointing	ρ(26) = 0.101, *p* = 0.866	ρ(31) = 0.038, *p* = 0.976	r(26) = 0.273, *p* = 0.152	r(31) = 0.057, *p* = 0.75
CCT	ρ(20) = 0.038, *p* = 0.984	ρ(26) = 0.436, *p* = 0.07	*r(20) = 0.491, p = 0.02*	r(26) = -0.177, *p* = 0.555
Consonants correct (NWR)	ρ(25) = 0.109, *p* = 0.852	ρ(29) = 0.276, p = 0.308	r(25) = 0.119, *p* = 0.554	r(29) = -0.12, *p* = 0.521
Syllables correct (NWR)	ρ(25) = 0.134, *p* = 0.785	ρ(29) = 0.205, p = 0.522	r(25) = 0.195, *p* = 0.329	r(29) = -0.021, *p* = 0.911
Correct stress (NWR)	ρ(25) = 0.234, *p* = 0.477	ρ(29) = 0.174, *p* = 0.627	r(25) = 0.043, *p* = 0.831	r(29) = -0.126, *p* = 0.5
Parent-estimated measures				
Receptive vocabulary (CDI)	ρ(24) = 0.25, *p* = 0.443	**ρ(26) = 0.536, *p* = 0.018**	r(24) = 0.199, *p* = 0.331	r(26) = -0.334, *p* = 0.083
Productive vocabulary (CDI)	ρ(24) = 0.102, *p* = 0.874	**ρ(26) = 0.547, *p* = 0.015**	r(24) = 0.232, *p* = 0.255	r(26) = -0.174, *p* = 0.376

1012 Bold font: $p < 0.01$ (infant-led) or $p < 0.025$ (parent-estimated). Italic: $p < 0.05$.

Table 5: All circular-linear (phase angle) and linear (vector length) correlations between AV-V phase angle/vector length over left temporal ROI, and language measures. Alpha is corrected for multiple comparisons and is equal to 0.01 for the infant-led and 0.025 for the parent-estimated measures, accounting for five different infant-led measures and two different parent-reported vocabulary measures.

	Circular-linear correlations with phase angle		Linear correlations with vector length	
	5/6 months	8/9 months	5/6 months	8/9 months
Infant-led measures				
Pointing	ρ(13) = 0.188, *p* = 0.766	ρ(19) = 0.424, *p* = 0.152	r(13) = 0.308, *p* = 0.247	r(16) = -0.006, *p* = 0.98
CCT	ρ(11) = 0.406, *p* = 0.343	*ρ(16) = 0.606, p = 0.037*	r(11) = 0.294, *p* = 0.33	r(14) = 0.332, *p* = 0.209
Consonants correct (NWR)	ρ(12) = 0.597, *p* = 0.08	ρ(18) = 0.125, *p* = 0.856	r(12) = 0.486, *p* = 0.078	r(15) = -0.073, *p* = 0.781
Syllables correct (NWR)	ρ(12) = 0.63, *p* = 0.062	ρ(18) = 0.046, *p* = 0.979	*r(12) = 0.557, p = 0.038*	r(15) = -0.045, *p* = 0.865
Correct stress (NWR)	ρ(12) = 0.629, *p* = 0.063	ρ(18) = 0.342, *p* = 0.31	*r(12) = 0.576, p = 0.031*	r(15) = -0.315, ***p*** = 0.218
Parent-estimated measures				
Receptive vocabulary (CDI)	ρ(12) = 0.521, *p* = 0.149	ρ(16) = 0.125, *p* = 0.869	r(12) = 0.286, *p* = 0.322	r(13) = 0.153, *p* = 0.587

Productive vocabulary (CDI)	ρ(12) = 0.587, *p* = 0.09	ρ(16) = 0.244, *p* = 0.586	r(12) = 0.342, *p* = 0.232	r(13) = 0.229, *p* = 0.411

Bold font: $p < 0.01$ (infant-led) or $p < 0.025$ (parent-estimated). Italic: $p < 0.05$.

Figure captions:

Figure 1: Relative power values for visual speech, resting state and AV-V conditions, across frequency bins. Relative power during the syllable stimulus is shown in blue and during the resting state in red. Panel (a) shows group mean relative power (shaded area indicates

standard error) for the five-month visual speech data; panel (b) for the eight-month visual speech data; panel (c) plots the difference in relative power between the AV speech at six months and the visual-only speech at five months; and panel (d) plots the difference in relative power between the AV speech at nine months and the visual-only speech at eight months. Analyses related to panels (a) and (b) are in section 3.1.1; those related to (c) and (d) are in 3.3.1.

Figure 2: Circular plots showing individual mean phase angles and vector length in blue and group means in yellow for the visual-only condition and the AV-V condition.

Figure 3: Relative power at 2Hz, plotted by age group and condition. The 5mo (five-month) and 8mo (8-month) data relates to the visual-only speech; the 6mo (six-month) and 9mo (nine-month) data is from the AV speech. Each dot is an individual infant's mean relative power at 2Hz during the resting state (red, left) or syllable stimulus (blue, right) (Outliers were already excluded for AV and VO separately, with no 2Hz values reaching outlier exclusion criteria, consequently no further exclusions were performed).

Figure 4: Significant relationships between phase angle or vector length in visual speech or AV-V analyses, and later language measures. Panel A depicts 2Hz phase angle in response to visual-only speech at 8 months and receptive vocabulary on the CDI; panel B depicts 2Hz phase angle in response to visual-only speech at 8 months and productive CDI vocabulary. In panels A and B, the blue dots are participants, ordered from the participant with the smallest phase angle to the largest (expressed in radians). Values are shown on the left y-axis. The red dots are language outcome measures, with values shown on the right y-axis. The red line is plotted by the geom_smooth function in ggplot2 using the loess method to illustrate the trend

1067 in the data points. The sinusoidal aspect of the trend lines is indicative of the underlying
1068 circular-linear relationship to the phase angle data.