The origins and development of speech envelope tracking during the first months of life

Maria Clemencia Ortiz Barajas [a],*, Ramón Guevara [b], Judit Gervain [a,c]

[a] *Integrative Neuroscience and Cognition Center, CNRS & Université de Paris, Paris, France*
[b] *Department of Physics and Astronomy, University of Padua, Padua, Italy*
[c] *Department of Developmental Psychology and Socialization, University of Padua, Padua, Italy*

ABSTRACT

When humans listen to speech, their neural activity tracks the slow amplitude fluctuations of the speech signal over time, known as the speech envelope. Studies suggest that the quality of this tracking is related to the quality of speech comprehension. However, a critical unanswered question is how envelope tracking arises and what role it plays in language development. Relatedly, its causal role in comprehension remains unclear, as some studies have found it to be present even for unintelligible speech. Using electroencephalography, we investigated whether the neural activity of newborns and 6-month-olds is able to track the speech envelope of familiar and unfamiliar languages in order to explore the developmental origins and functional role of envelope tracking. Our results show that amplitude and phase tracking take place at birth for familiar and unfamiliar languages alike, i.e. independently of prenatal experience. However, by 6 months language familiarity modulates the ability to track the amplitude of the speech envelope, while phase tracking continues to be universal. Our findings support the hypothesis that amplitude and phase tracking could represent two different neural mechanisms of oscillatory synchronisation and may thus play different roles in speech perception.

1. Introduction

How does our brain handle the challenging task of processing the rapidly unfolding speech signal? The exact mechanisms the human brain implements to decode speech are not yet fully understood. One neural process that has been claimed to play a role is the tracking of the speech envelope. The speech envelope, corresponding to the slow overall amplitude fluctuations of the speech signal over time, with peaks occurring roughly at the syllabic rate, has been argued to play a crucial role in speech perception. Behaviourally, comprehension is impaired when the speech envelope is suppressed (Drullman et al., 1994a, b), while adult listeners readily understand degraded speech in which only the envelope is preserved, at least when speech is presented in silence (Shannon et al., 1995). Additionally, neuroimaging studies have shown that when adults listen to speech, their brains synchronize with specific features of the speech envelope, a phenomenon known as speech envelope tracking (Ahissar et al., 2001; Luo and Poeppel, 2007; Abrams et al., 2008; Nourski et al., 2009). One feature with which brain activity may synchronize is the amplitude of the speech signal: amplitude synchronization occurs when the amplitude of the neural activity in the auditory cortex follows the contours of the speech envelope (Abrams et al., 2008; Nourski et al., 2009; Kubanek et al., 2013). The auditory cortex also synchronizes with the phase of the speech envelope by modulating the phase of its ongoing oscillations to match the phase of the envelope (i.e. phase-locking) (Peelle et al., 2013; Pefkou et al., 2017). The quality of amplitude (Ahissar et al., 2001; Nourski et al., 2009) and phase synchronization (Luo and Poeppel, 2007; Peelle et al., 2013) has been found to correlate with comprehension. These findings led to the conclusion that speech envelope tracking is a key mechanism in speech comprehension.

Recent electrophysiological evidence, however, has questioned this proposal, as phase-locking to the speech envelope in the theta band (4−8 Hz), the oscillation whose frequency corresponds to the modulation frequency of the speech envelope, i.e. approximately the syllabic rate (4−5 Hz) (Ding et al., 2017; Varnet et al., 2017), has been found to be independent of comprehension. Brain responses in the theta band have been shown to track the speech envelope even when speech is time-compressed at a rate that renders it incomprehensible to adult listeners (Pefkou et al., 2017; Zoefel and VanRullen, 2016). Moreover, normal sentences, readily comprehensible to listeners, have been shown

---

to evoke similar phase synchronization as their unintelligible, time-reversed counterparts (Howard and Poeppel, 2010); and stronger phase-locking results have been found in adults when listening to sentences in their second language than in their native language (Song and Iverson, 2018). These results thus call into question whether envelope tracking is sufficient for comprehension to take place. More generally, it remains controversial whether speech comprehension and envelope tracking are causally linked (Zoefel et al., 2018).

Understanding how envelope tracking emerges during development offers unique insight into this question, as human infants' speech perception abilities are sophisticated already at birth, while comprehension doesn't arise until later in development. We, therefore, investigated infants' speech envelope tracking ability at birth and at 6 months of age in familiar and unfamiliar languages. This allowed us to address: (i) whether envelope tracking is present even in the absence of language comprehension, i.e. whether it is a basic auditory ability, or whether knowledge of language and thus comprehension are necessary for it to emerge, and (ii) whether familiarity with the language modulates it.

This is the first study to assess this neural ability at birth. In older infants, two existing studies have investigated envelope tracking, with a different focus from the questions we ask here. One study (Ortiz-Mantilla et al., 2013) found phase-locking to both native and non-native syllabic contrasts in the 2−4 Hz range in the left and right auditory cortices, as well as in the 3−5 Hz range in the anterior cingulate cortex. This phase-locking was found in frequency ranges that largely overlap with the modulation frequency of the speech envelope (4−5 Hz), possibly reflecting phase tracking of the speech envelope at 6 months. Another study investigated whether 7-month-olds track the speech envelope differently in infant-directed speech (IDS) and adult-directed speech (ADS) in their native language (Kalashnikova et al., 2018). They found no amplitude tracking for ADS and localized tracking (in the frontal area) for IDS. These studies suggest that envelope tracking may play a role in certain aspects of speech processing, such as phoneme discrimination or attention to IDS by 6–7 months of age. Our study aimed to investigate the role of experience in the emergence of envelope tracking.

We tested 47 full-term, healthy newborns, born to French monolingual mothers, within their first 5 days of life. Their experience with speech was, therefore, mostly prenatal. Prenatal speech experience consists of a signal that is filtered by the maternal tissues and the amniotic fluid, suppressing individual speech sounds, but preserving speech rhythm and melody (Vince and Armitage, 1980; Gerhardt et al., 1990; Lecanuet and Schaal, 2002), i.e. the speech envelope. We tested newborns with naturally spoken sentences in three languages: their native language, i.e. the language heard prenatally (French), a rhythmically similar unfamiliar language (Spanish), and a rhythmically different unfamiliar language (English). If envelope tracking (either amplitude or phase tracking) is responsible for language comprehension, we should not find it in newborns, who do not yet know the vocabulary and the grammar of their native language and thus do not comprehend it, despite prenatal familiarity with its sound pattern. By contrast, if envelope tracking is a basic auditory ability prior to language comprehension, newborns may already show it. Further, if prenatal experience shapes envelope tracking, then we might find differences between the three languages, since behavioural studies have shown that newborns are able to recognize their native language (Mehler et al., 1988; Moon et al., 1993), and to discriminate languages that are rhythmically different, but not those that are rhythmically similar (Nazzi et al., 1998; Ramus et al., 2000).

Specifically to our study, there may be an advantage for French, the language heard prenatally, and this advantage may also extend to Spanish, the rhythmically similar unfamiliar language, as newborns do not discriminate rhythmically similar languages behaviourally, and consequently their neural responses could be similar for the two languages due to their rhythmic closeness. By contrast, newborns can behaviourally discriminate French from English, the rhythmically different unfamiliar language, therefore their neural responses to these two languages may be different. These differences in neural processing, i.e. due to rhythmic differences and prenatal experience, could modulate envelope tracking.

To better understand the role of envelope tracking in language comprehension, it is also crucial to investigate whether envelope tracking changes with infants' growing knowledge of language over development. To address this question, we also tested 25 full-term, healthy 6-month-old French learning infants on the same three language conditions that we tested at birth, i.e. French, Spanish and English. By 6 months of age, infants can discriminate even rhythmically similar languages, as long as one of them is familiar to them (Bosch and Sebastián-Gallés, 1997; Molnar et al., 2013). Furthermore, by 6 months, infants begin to have considerable knowledge of their native language. They start to attune to the sound patterns of their native language, especially to vowels (Kuhl et al., 1992; Tsuji and Cristia, 2014), they start to learn the first word forms (Tincoff and Jusczyk, 1999; Bergelson and Swingley, 2012) and begin to acquire some basic grammatical properties of their native language, such as word order (Gervain et al., 2008; Gervain and Werker, 2013).

If experience with speech or maturation facilitates tracking generally, then we should find enhanced synchronization (in either amplitude or phase) at 6 months compared to birth for all three languages. Alternatively or additionally, if familiarity with the native language preferentially modulates envelope tracking, then we might find synchronization to the native language (French) to be different from that of the unfamiliar languages (Spanish and English). It may also be the case that emerging linguistic knowledge such as attunement to the native phoneme repertoire, word learning or the acquisition of grammar may enhance the relevance of the finer details of speech such as individual phonemes, the acoustic correlates of which are not present in the envelope, but in the fine structure of speech (Rosen, 1992). Indeed, the prosody of the native language is already heard in utero (Vince and Armitage, 1980; Gerhardt et al., 1990; Lecanuet and Schaal, 2002) and already shapes newborns' speech perception (Abboub et al., 2016) and cry production abilities (Mampe et al., 2009). At birth, suprasegmental units such as syllables and prosodic contours often carry more weight in for infants than phonemes (Bertoncini and Mehler, 1981; Bertoncini et al., 1995; Mehler et al., 1996; Benavides-Varela and Gervain, 2017). After several months of postnatal experience with the native language, infants start to narrow their initially universal and broad-based perceptual abilities down the phoneme inventory of the native language (Werker and Tees, 1984; Kuhl et al., 1992), losing the ability to discriminate non-native contrasts, while sharpening their native phoneme categories. In parallel, they begin to extract the first word forms from speech and associate them to objects (Tincoff and Jusczyk, 1999; Bergelson and Swingley, 2012) and start to learn the basic word order of their native language (Gervain et al., 2008). These abilities all require infants to pay attention to the finest details of individual phonemes, and at this age, they are even sensitive to within-category phonetic differences (McMurray and Aslin, 2005). We thus hypothesize that these developmental changes temporarily reduce or attenuate envelope tracking (either amplitude or phase tracking), which mainly focuses on the syllabic level. Since these developmental changes only occur for the native language, such a reduction is expected to impact the native language more than the unfamiliar languages.

2. Materials and methods

2.1. Participants

The protocol for this study was approved by the CER Paris Descartes ethics committee of the Paris Descartes University (current Université de Paris). All parents gave written informed consent prior to participation, and were present during the testing session.

2.1.1. Newborns

We recruited newborn participants at the maternity ward of the Robert-Debré Hospital in Paris, and we tested them during their hospital stay. The inclusion criteria for our newborn group were: i) being full-term and healthy, ii) having a birth weight > 2800 g, iii) having an Apgar score > 8, iv) being maximum 5 days old, and v) being born to French native speaker mothers who spoke this language at least 80 % of the time during the last trimester of the pregnancy according to self-report. We tested a total of 55 newborn participants, and excluded 8 participants from data analysis due to: bad data quality resulting in an insufficient number of non-rejected trials (n = 4), not finishing the experiment due to fussiness and crying (n = 3), or technical problems (n = 1). Thus, electrophysiological data from 47 newborns (age 2.45 days ± 1.18 d; range 1–5 d; 20 girls, 27 boys) were included in the analysis.

2.1.2. Six-month-olds

We recruited 6-month-old infants in two different ways: we re-contacted the participants who took part of the newborn experiment and 36 % of them came back (n = 20, longitudinal), and we recruited new participants through the database of our Babylab (INCC – Université de Paris) who were tested for the very first time (n = 14, cross-sectional). The inclusion criteria were: i) being full-term and healthy, ii) being 6 months old ± 15 days, and iii) being exposed to French at least 80 % of the time since birth (according to parental report). Due to availability constraints of the re-contacted families, we tested 4 participants outside of the intended age range: one 5-month-old, one 10-month-old, and two 11-month-olds. We tested a total of 34 infants, and excluded 9 participants from data analysis due to: bad data quality resulting in an insufficient number of non-rejected trials (n = 4), not finishing the experiment due to crying (n = 2), or technical problems (n = 3). Thus, electrophysiological data from 25 infants (age 204 days ± 48 d; range 152–340 d; 10 girls, 15 boys) were included in the analysis (12 longitudinal, 13 cross-sectional).

2.2. Procedure

We tested infants by presenting them with naturally spoken sentences in three languages while simultaneously recording their neural activity using electroencephalography (EEG).

2.2.1. Newborns

The EEG recordings were conducted in a dimmed, quiet room at the Robert-Debré Hospital in Paris. Participants were divided into 3 groups, and each group heard one of three possible sets of sentences (*SI Appendix*, Table S1): 14 newborns heard set1, 16 newborns heard set2, and 17 newborns heard set3. During the recording session, newborns were comfortably asleep or at rest in their hospital bassinets (Fig. 1a). The stimuli were delivered bilaterally through two loudspeakers positioned on each side of the bassinet using the experimental software E-Prime. The sound volume was set to a comfortable conversational level (~65−70 dB). Following the procedure of a previous study investigating envelope tracking in older children (Abrams et al., 2008), we presented participants with one sentence per language, and repeated it 100 times to ensure sufficiently good data quality. The experiment consisted of 3 blocks, each block containing the 100 repetitions of the test sentence in a given language, each block thus lasted around 7 min. An interstimulus interval of random duration (between 1–1.5 s) was introduced between sentence repetitions, and an interblock interval of 10 s was introduced between language blocks (Fig. 1c). The order of the languages was pseudo-randomized and approximately counterbalanced across participants. The entire recording session lasted about 21 min.

2.2.2. Six-month-olds

The EEG recordings were conducted at two different locations in order to accommodate the constraints of the re-contacted families: one was a dimmed, quiet room at the Robert-Debré Hospital in Paris (n =
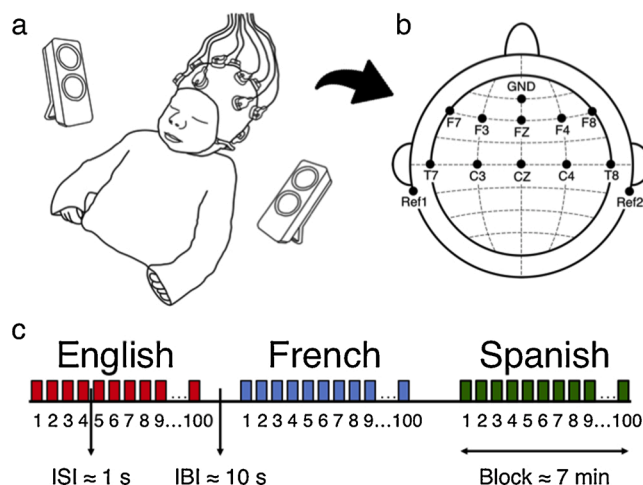

Fig. 1. EEG experimental setup and design. (a) Newborn with EEG cap. (b) Location of recorded channels according to the international 10-20 system. (c) Experiment block design. ISI: Interstimulus interval, IBI: Interblock interval.

10); and the other one was a sound-attenuated testing booth at the Babylab of the Université Paris Descartes (n = 15). The setups were identical in the two locations. Participants heard one of three possible sets of sentences (*SI Appendix*, Table S1): 10 infants heard set1, 14 infants heard set2, and 1 infant heard set3. For the participants who were also tested at birth, we controlled for the set of sentences that was presented to them at 6 months, to ensure that they did not hear the same stimuli and rule out the effect of any memory trace. During the recording session, infants were seated on a parent's lap. In order to avoid motion artifacts and maintain infants' attention, we also presented a silent video of a children's animated cartoon on a monitor placed approximately 60 cm in front of them during the entire recording session. The stimuli were delivered bilaterally through two loudspeakers positioned on each side of the monitor using the experimental software E-Prime. The sound volume was set to a comfortable conversational level (~65−70 dB). We presented participants with one sentence per language, and repeated it 50 times to ensure sufficiently good data quality. For this experiment we used fewer repetitions than for the newborn group to reduce testing time as is appropriate for the attention span of 6-month-old infants. The experiment consisted of 3 blocks, each block containing the 50 repetitions of the test sentence in a given language, each block thus lasted around 3.5 min. An interstimulus interval of random duration (between 1–1.5 s) was introduced between sentence repetitions, and an interblock interval of 10 s was introduced between language blocks. The entire recording session lasted about 11 min. When investigating envelope tracking at 6 months, we had a special interest in the rhythmically similar languages (French and Spanish), because at this age infants can discriminate them (Bosch and Sebastián-Gallés, 1997; Molnar et al., 2013), whereas they could not do it at birth (Nazzi et al., 1998; Ramus et al., 2000). For this reason, we presented French and Spanish in the first two blocks, to maximize the chances that we could collect data for these conditions. However, the order of the first two blocks (French and Spanish) was pseudo-randomized and counterbalanced across participants, and English was always presented in the third block. We found no order effects related to this design (*SI Appendix*, Table S10 and Table S13).

2.3. Stimuli

We tested infants in the following three languages: their native language (French), a rhythmically similar unfamiliar language (Spanish), and a rhythmically different unfamiliar language (English). The stimuli consisted of sentences taken from the story Goldilocks and the Three Bears. Three sets of sentences were used, where each set comprised the

translation of a single utterance into the 3 languages (English, French and Spanish). The translations were slightly modified in order to match sentence duration across languages within the same set (see *SI Appendix*, Table S1). All sentences were recorded in mild infant-directed speech by a female native speaker of each language (a different speaker for each language), at a sampling rate of 44.1 kHz. There were no significant differences between the sentences in the three languages in terms of minimum and maximum pitch, pitch range and average pitch (see *SI Appendix*, Table S1). We computed the amplitude and frequency modulation spectra of the sentences in the three languages as defined by Varnet et al. (2017) to explore if utterances were consistently different across languages (see *SI Appendix*, Fig. S1). We found that utterances were similar in every spectral decomposition. The intensity of all recordings was adjusted to 77 dB. We used the same stimuli when testing newborns and 6-month-olds.

2.4. Data acquisition

We recorded the EEG data with active electrodes and an acquisition system from Brain Products (Hardware: actiCAP and actiCHamp, Brain Products GmbH, Gilching, Germany). We used a 10-channel layout to acquire cortical responses from the following scalp positions: F7, F3, FZ, F4, F8, T7, C3, CZ, C4, T8 (Fig. 1b). We chose these recording locations in order to include those where auditory and speech perception related neural responses are typically observed in infants (Stefanics et al., 2009; Tóth et al., 2017; channels T7 and T8 used to be called T3 and T4 respectively). We used two additional electrodes placed on each mastoid for online reference, and a ground electrode placed on the forehead. We recorded the data at a sampling rate of 500 Hz. The electrode impedances were kept below 140 kΩ for newborns and 40 kΩ for 6-month-olds.

2.5. Amplitude tracking analysis

We pre-processed and analysed all the data using custom Matlab® scripts. We implemented the same data processing and analysis pipeline for the two groups of participants (newborns and 6-month-olds).

2.5.1. EEG processing

We band-pass filtered the continuous EEG data between 1 and 40 Hz using a zero phase-shift Chebyshev filter, and then segmented it into a series of 2,960-ms long epochs. Each epoch started 200 ms before the utterance onset (corresponding to the pre-stimulus baseline), and contained a 2,760 ms long post-stimulus interval. We baseline corrected all epochs using their 200 ms pre-stimulus period, and then submitted them to a three-stage rejection process to exclude the contaminated ones. First, we rejected epochs with amplitude exceeding ±75 μV. Second, we rejected those whose standard deviation (SD) was higher than 3 times the mean SD of all non-rejected epochs, or lower than one-third the mean SD. Third, we visually inspected the remaining epochs to remove any residual artifacts. After epoch rejection was completed, we computed the evoked response by averaging all remaining epochs. This averaging was done per participant, per language and per channel. It needs to be noted that evoked responses were extracted from the complete EEG signal (band-pass filtered, 1-40 Hz), and not from its envelope. At birth, participants contributed on average 49 epochs (SD: 12.6; range across participants: 30–82) for English, 52 epochs (SD: 15.3; range across participants: 28–88) for French, and 49 epochs (SD: 13.5; range across participants: 22–76) for Spanish. Participants who had less than 20 remaining epochs after epoch rejection (at birth) were not included in the data analysis (n = 4, as also reported in the Participants section). At six months, participants contributed on average 27 epochs (SD: 8.2; range across participants: 11–44) for English, 31 epochs (SD: 10.1; range across participants: 11–49) for French, and 31 epochs (SD: 9.1; range across participants: 13–47) for Spanish. Participants who had less than 10 remaining epochs after epoch rejection (at six months) were not

included in the data analysis (n = 4, as also reported in the Participants section).

2.5.2. Stimuli processing

The amplitude envelopes of the speech signals were obtained by calculating the magnitude of the Hilbert transform of the stimulus waveforms. The envelopes were then low-pass filtered at 40 Hz and down-sampled to 500 Hz to match the characteristics of the evoked responses.

2.5.3. Amplitude tracking assessment

We assessed amplitude tracking by performing cross-correlation analysis between the speech envelopes and the evoked responses at the participant level. To implement the cross-correlation analysis, we computed the Spearman correlation between the two signals for lags between 80 ms and 600 ms (in 2 ms steps). At each step, we calculated the correlation for the "envelope-following" period (defined as the time range after 250 ms in Abrams et al., 2008) of both signals. The time range from 0 to 250 ms was not included in the correlation as it was considered to be an onset period (as in Abrams et al., 2008), where large-amplitude early auditory responses occur. The latency of the peak of the cross-correlation function corresponds to the time lag at which the EEG response and the stimulus envelope are most similar, i.e. the time lag at which amplitude tracking takes place. We performed the cross-correlation analysis per participant, per language and per channel. To ensure that an optimal lag could be found, we averaged the cross-correlation functions across the channels that were at least weakly correlated ($\rho > 0.15$), as done by Kubanek and colleagues (2013), to obtain the mean cross-correlation function for each language at the participant level (*SI Appendix*, Fig. S2). We then computed a time lag for each language at the participant level (*SI Appendix*, Fig. S3) as the latency of the highest peak in the mean cross-correlation function of the given language. We submitted the time lags from all the participants to a repeated measures ANOVA with *Language* (French/Spanish/English) as a within-subjects factor, and it yielded no significant main effect of *Language* on the time lags [F(2,92) = 0.044, p = 0.957]. It is important to note that the mean cross-correlation functions from some of the languages at the participant level did not exhibit well defined peaks (*SI Appendix*, Fig. S2b) that would allow the identification of an optimal time lag for the given language. Therefore, we computed a mean time lag for each participant by averaging the mean cross-correlation functions from the languages that were at least weakly correlated ($\rho > 0.15$). As a last step, we quantified amplitude tracking as the correlation between the evoked responses and the speech envelopes evaluated at the mean time lag obtained for each participant. The correlation values were Fisher transformed before statistical analysis as is customary when performing statistical comparisons of correlation coefficients. Even though only weakly correlated channels were included when obtaining the participants' time lags, all channels were included in the amplitude tracking correlation analysis.

2.5.4. Habituation effect assessment

To investigate if stimulus repetition introduced habituation that could have had an effect on our correlation results, we assessed amplitude tracking at the single trial level on all the non-rejected trials for each participant separately. To do so, we calculated the correlation between the speech envelope and the single trial EEG responses. The correlation values were Fisher transformed before statistical analysis. We submitted the correlation values from each language and channel to simple linear regression analyses with *Trial* as the predictor variable to evaluate the effect of habituation. The obtained p-values were corrected for multiple comparisons applying the False Discovery Rate method (Benjamini and Hochberg, 1995).

2.6. Phase tracking analysis

We pre-processed and analysed all the data using custom Matlab® scripts. We implemented the same data processing and analysis pipeline for the two groups of participants (newborns and 6-month-olds). To assess phase tracking, we first determined the syllable rate of each sentence by detecting the number of peaks per second in their speech envelopes, and rounded the values to the closest integer (as performed by Pefkou et al., 2017). The syllabic rate of the tested sentences ranged from 3 to 6 Hz (see *SI Appendix*, Table S1).

2.6.1. EEG processing

We band-pass filtered the continuous EEG data between 3 and 6 Hz using a zero phase-shift Chebyshev filter. After filtering, we submitted the EEG data to the same processing steps as described in the *Amplitude tracking analysis* section: epoching, baseline correction, epoch rejection, and averaging of remaining epochs to obtain the evoked response. The epochs rejected from the EEG data filtered at the syllabic rate were those identified as contaminated by the three-stage process run on the *Amplitude tracking analysis*.

2.6.2. Stimuli processing

To match the characteristics of the filtered EEG, the speech envelopes were also band-pass filtered between 3 and 6 Hz.

2.6.3. Phase tracking assessment

We assessed phase tracking by computing the phase-coherence between the filtered evoked responses and the filtered speech envelopes by following these steps: 1) obtaining the Hilbert transform of both signals, 2) calculating the unitary signals of each analytic signal, and 3) computing the phase synchronization between both unitary signals. This analysis was performed per participant, per language and per channel.

2.7. Statistical analysis

To validate our results and establish envelope tracking, we assessed our correlation and phase-coherence results to determine if their distributions were significantly different from zero and from chance (permutation test). Furthermore, we implemented repeated measures ANOVAs to evaluate whether our envelope tracking results varied across languages and/or channels.

2.7.1. Comparison to zero

To assess whether the correlation and phase-coherence distributions were significantly different from zero, we submitted them to one-sample t-tests (two-tailed). This was done for each language and each channel. The p-values from these t-tests were corrected for multiple comparisons (i.e. to adjust the family-wise error rate for each comparison) using the Bonferroni correction.

2.7.2. Permutation test

To assess whether the correlation and phase-coherence distributions were significantly different from chance, we submitted them to a permutation test. To do so, we shuffled the EEG data 1000 times along the time dimension while keeping the other labels intact (participant, language, and channel) (Hurtado et al., 2004; Sun et al., 2011). At each iteration we performed the amplitude and phase tracking analyses in the same way as done on the non-permuted data. For the amplitude tracking analysis this involved: 1) running the cross-correlation analysis between the permuted EEG and the speech envelopes, 2) computing the time lags, 3) obtaining the correlation values at the given time lags, and 4) Fisher transforming the correlation values. And for the phase tracking analysis this involved computing the phase-coherence between the permuted EEG and the speech envelopes. The correlation and phase-coherence distributions obtained from the permuted data served as a baseline of comparison for our amplitude and phase tracking results. The p-values

for the permutation test were derived as the number of permutations out of the 1000 whose correlation/phase-coherence distributions were not significantly different from the distributions obtained for the non-permuted data (in paired samples t-tests).

2.7.3. ANOVAs

To assess whether amplitude and phase tracking were different across languages and/or channels, we submitted the Correlation and Phase-coherence values, from each age group, to repeated measures ANOVAs in DataDesk, with *Language* (French/Spanish/English) and *Channel* (F7 / F3 / Fz / F4 / F8 / T7 / C3 / Cz / C4 / T8) as the within-subjects factors, and *Participant* as a random factor. This analysis was performed on the results from the channels that significantly tracked at least one of the languages; therefore, channel T8, which did not show envelope tracking for any language (see Section 3.1.1 below), was excluded from the ANOVA on amplitude tracking results at birth.

It was not possible to compare the two ages directly, as a number of reasons preclude valid statistical comparisons: most importantly, newborns and 6-month-olds were not in the same state of alertness, furthermore, the sample sizes and the number of included trials, and hence the signal-to-noise ratios also differed between the two age groups.

3. Results

3.1. Speech envelope tracking at birth

3.1.1. Amplitude tracking

We assessed amplitude synchronization by performing cross-correlation analysis between the speech envelopes and the neural responses at the participant level. The peak in the cross-correlation function corresponds to the time lag at which the two signals are most similar, i.e. the time lag at which envelope tracking takes place. We found that the strongest correlations occurred at similar time lags for the three languages: mean time lag for English: 293 ms, for French 282 ms, and for Spanish 308 ms (*SI Appendix*, Fig. S3). The time lags were not significantly different across languages (repeated measures ANOVA F $(2,92) = 0.044$, p $= 0.957$), so we calculated a mean time lag for each participant (see Materials and Methods), and used it to obtain an amplitude synchrony measure, i.e. correlation, for each participant, channel, and language.

Fig. 2 illustrates the envelope tracking results of a single participant, showing how the EEG recording (blue curves) follows the speech envelope (black curves) for a French sentence at each channel. Mean correlation values across the whole group of infants were 0.099 (SD: 0.0240; range across channels: 0.048−0.128) for English, 0.122 (SD: 0.0191; range across channels: 0.090−0.142) for French, 0.118 (SD: 0.0273; range across channels: 0.059−0.152) for Spanish (see *SI Appendix*, Table S2). These correlations are of a similar magnitude as those observed in adults tested with ECoG (electrocorticography; mean correlation values ranged between 0.05 and 0.20) (Kubanek et al., 2013), despite the fact that ECoG is a more sensitive technique than scalp EEG.

To establish whether these correlation values do indeed indicate amplitude tracking, we assessed if their distributions were significantly different from zero, with one-sample t-tests (two-tailed) (Kubanek et al., 2013), and from chance, with a permutation test (Peelle et al., 2013) (*SI Appendix*, Table S2, Table S3). We permuted the EEG data 1000 times, and performed the envelope tracking analysis at each permutation. The correlation values were Fisher transformed before statistical analysis. The amplitude tracking results (Fig. 3a) from the measured EEG data were significant, defined as higher than zero and different from chance in the permutation test, for every channel and language, with the exception (Fig. 3a) of T7 for French and English, and T8 for all three languages. These results thus establish that the newborn brain tracks the amplitude of the speech envelope in all three languages.

To assess possible differences across languages and channels, we
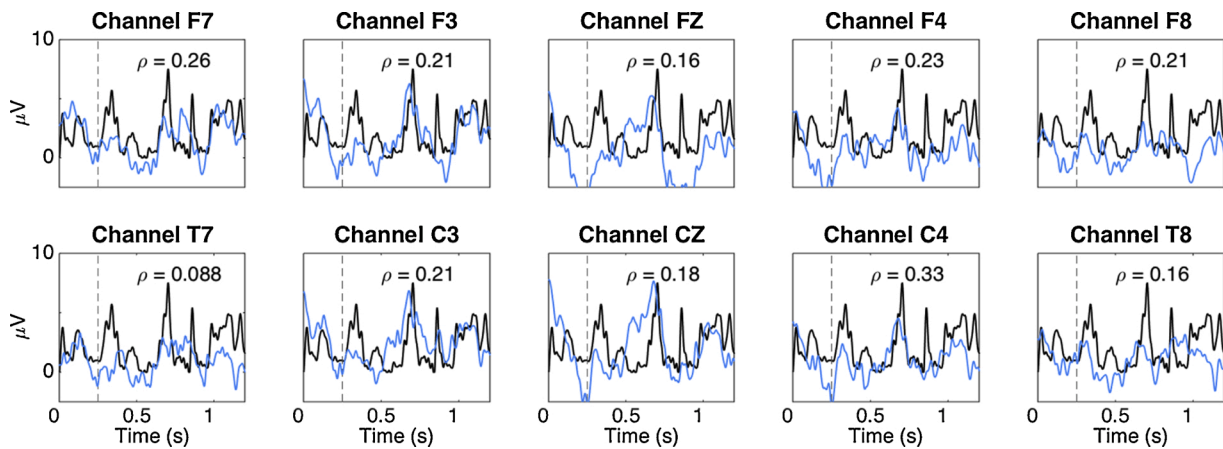
Fig. 2. Amplitude envelope tracking of a French sentence in a newborn participant. The black curves represent the speech envelope of a French sentence and the blue curves the cortical activity measured per channel. For visualization purposes the EEG data was shifted backward in time using the time lag between the neural response and the speech signal. The Spearman correlation between the two signals is indicated by "ρ". The vertical line at 250 ms indicates the beginning of the envelope-following period. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article).
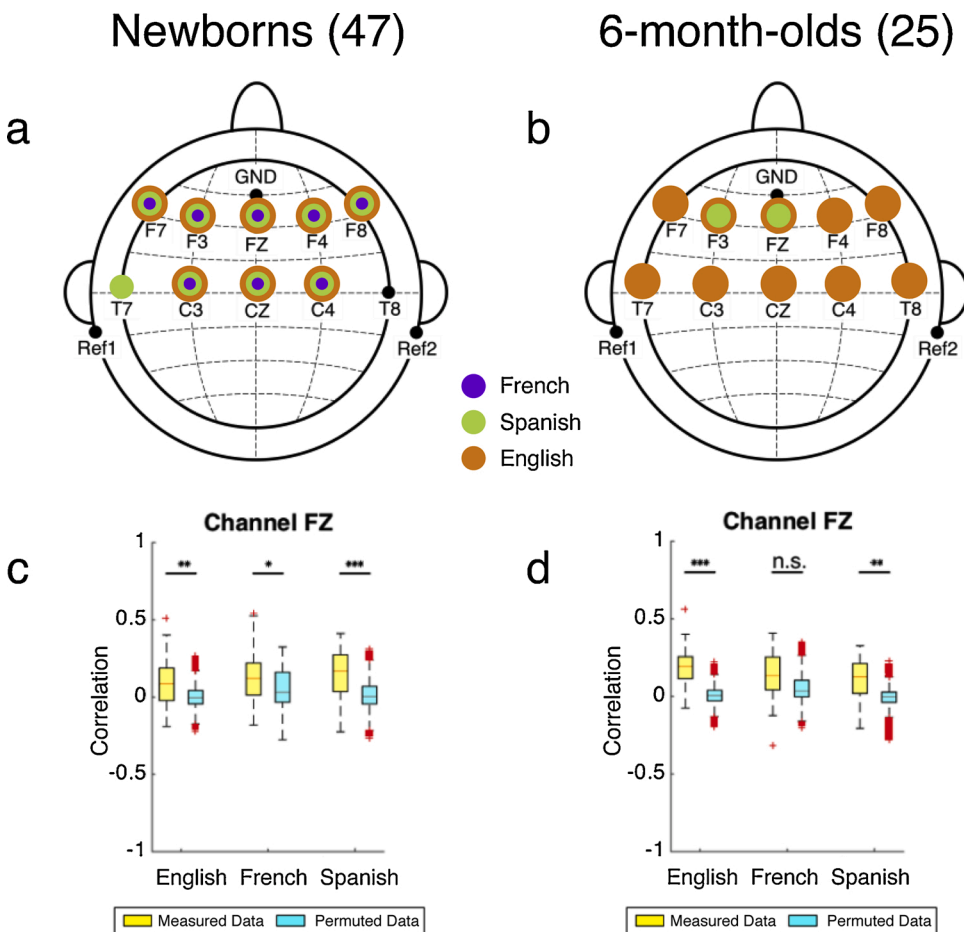

Fig. 3. Amplitude tracking results. (a, b) Topographic distribution of channels with significant (different from zero and from chance permutations) amplitude tracking results for each language (purple for French, green for Spanish, orange for English) for the (a) group of 47 newborns, and (b) group of 25 six-month-olds. (c, d) Amplitude tracking results for the three languages at channel Fz, as an illustration. The yellow boxes represent the distributions of the results from the measured EEG data, and the blue boxes represent the distributions of the results from the permuted data, for the (c) newborn group, and (d) 6-month-old group. n. s.: non-significant, * p < 0.05, ** p < 0.01, *** p < 0.001. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article).

submitted the correlation values from the 9 channels that were found to show tracking in any of the languages in the permutation analysis to a repeated measures ANOVA with *Language* (French/Spanish/English) and *Channel* (F7 / F3 / Fz / F4 / F8 / T7 / C3 / Cz / C4) as within-subjects factors. This ANOVA yielded no significant main effect of *Language* [$F_{(2,1196)}$ = 2.563, p = 0.078] or *Channel* [$F_{(8,1196)}$ = 1.152, p = 0.326]. These results imply that the newborn brain tracks the amplitude of the speech envelope equally well in the three languages in all nine

channels.

Additionally, we also investigated if habituation to the sentences due to the repetitions had an effect on our amplitude tracking results. Decreased responses due to habituation may hide differences across channels or languages. To test this, we computed amplitude tracking for single non-rejected trials and submitted their correlation values to linear regression analyses investigating the potential effects of habituation over the trials in each channel for each language. These analyses showed

that trial number did not significantly predict amplitude tracking (all p > 0.05, $R^2 < 0.003$) (*SI Appendix*, Table S4). Habituation effects thus did not impact the amplitude tracking results.

3.1.2. Phase tracking

To assess phase synchronization, we first determined the syllable rate of each sentence by detecting the number of peaks per second present in their speech envelopes, and rounded them to the closest integer (Pefkou et al., 2017). This resulted in a range of syllable rates from 3 to 6 Hz (*SI Appendix*, Table S1). We used the frequencies of the syllable rates to band-pass filter the EEG data and the speech envelopes, and extracted the phase of both filtered signals to compute the phase-coherence between them. Mean phase-coherence values across the whole group of infants were 0.319 (SD: 0.0200; range across channels: 0.292−0.350) for English, 0.331 (SD: 0.0147; range across channels: 0.302−0.349) for French, and 0.343 (SD: 0.0198; range across channels: 0.312−0.377) for Spanish (see *SI Appendix*, Table S5). Our results are numerically similar to what has been observed in adults (Pefkou et al., 2017), whose mean phase-coherence values were approximately 0.27.

To establish phase tracking, we assessed if the phase-coherence distributions were significantly different from zero (one-sample t-tests, two-tailed), and from chance (permutation test) (*SI Appendix*, Table S5, Table S6). We conducted a permutation analysis, similar to the one implemented for the amplitude tracking results. Fig. 4a indicates the channels where phase-coherence results were significantly higher than zero and chance. The phase tracking results from the measured data were significant (higher than zero and the chance permutation results) for every channel and language.

To assess possible differences in the phase synchronization results across languages and channels, we submitted the phase-coherence results from the 10 channels to a repeated measures ANOVA with *Language* and *Channel* as within-subjects factors. This ANOVA yielded no significant main effect of *Language* [$F(2,1334) = 2.744$, $p = 0.065$], or *Channel* [$F(9,1334) = 0.624$, $p = 0.777$]. These results imply that the newborn brain tracks the phase of the speech envelope equally well in the three languages in all channels. We did not investigate if habituation to the sentences due to the repetitions had an effect on our phase tracking results, because phase is not affected by the decreasing amplitude of the responses.

3.2. Speech envelope tracking at 6 months

3.2.1. Amplitude tracking

To assess amplitude synchronization we followed the same procedure as the one described for newborns. Mean correlation values across the whole group of infants were 0.163 (SD: 0.0324; range across channels: 0.094−0.217) for English, 0.114 (SD: 0.0195; range across channels: 0.082−0.149) for French, and 0.089 (SD: 0.0195; range across channels: 0.068−0.130) for Spanish (see *SI Appendix*, Table S7). These correlations are of a similar magnitude to those observed in newborns above, and in adults tested with ECoG (Kubanek et al., 2013).

To establish amplitude tracking, we assessed if the correlation distributions were significantly different from zero (one-sample t-tests, two-tailed), and from chance (permutation test), as done for the newborn group (*SI Appendix*, Table S7, Table S8). The correlation values were Fisher transformed before statistical analysis. Fig. 3b indicates the
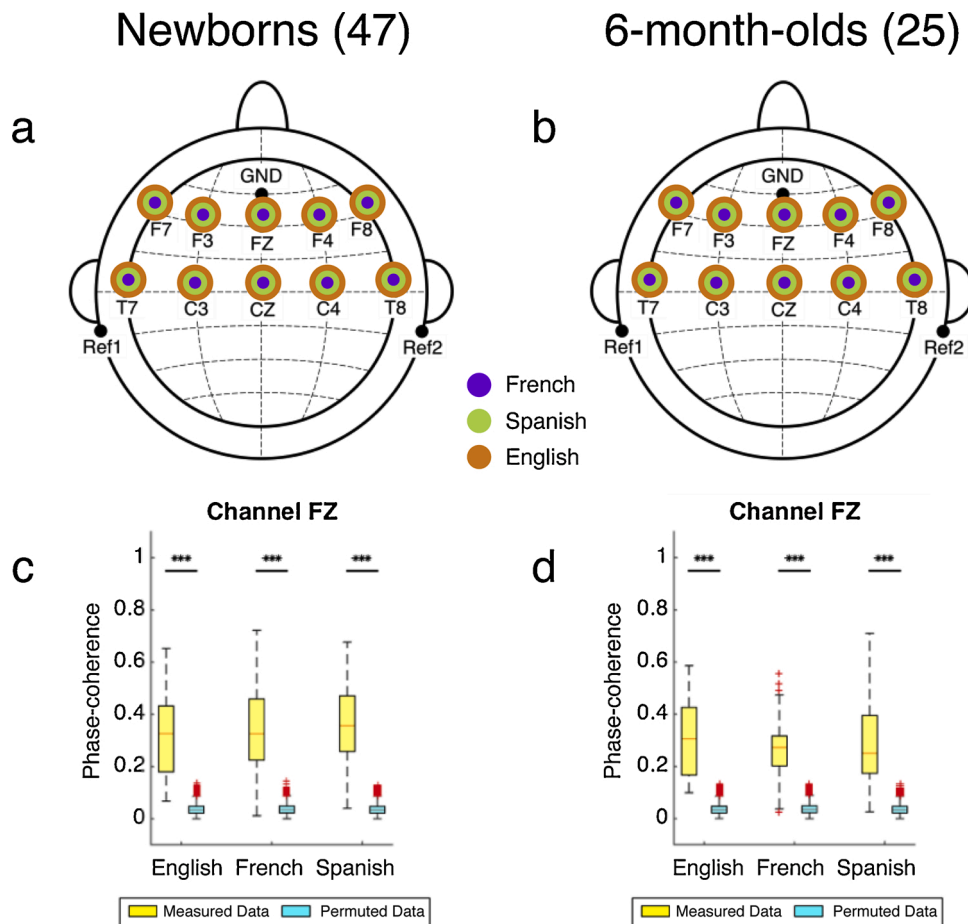

Fig. 4. Phase tracking results. (a, b) Topographic distribution of channels with significant phase tracking results for each language for the (a) newborn group, and (b) 6-month-old group. (c, d) Phase tracking results for the three languages at channel Fz, as an illustration, for the (c) newborn group, and (d) 6-month-old group. Plotting conventions as before.

channels where correlations were significantly higher than zero and chance. We found significant amplitude tracking only for English in all channels. For French, none of the channels showed significant amplitude tracking, while for Spanish channels F3 and Fz exhibited significant results.

To directly assess differences across languages and channels, we submitted the correlation values from the 10 channels to a repeated measures ANOVA with *Language* and *Channel* as within-subjects factors. This ANOVA yielded a highly significant main effect of *Language* [F (2,696) = 16.872, p < 0.0001, η^2 = 0.046], mainly driven by the significant differences between English and French (p < 0.001), and English and Spanish (p < 0.0001), with the difference between French and Spanish being non-significant (p = 0.063), as shown by LSD post hoc tests. No main effect of *Channel* was found [F(9,696) = 1.083, p = 0.373]. The *Language* effect is in line with the differences found across languages in the t-tests against zero and the permutation tests: our results show that the brain of 6-month-olds tracks the amplitude of the speech envelope for a rhythmically different unfamiliar language (English), but not for the native language (French), and only very weakly for a rhythmically similar unfamiliar language (Spanish).

We also investigated if habituation due to stimulus repetition had an effect on our amplitude tracking results at 6 months by following the same procedure as the one described for newborns. We found that trial number did not significantly predict amplitude tracking (all p > 0.05, R^2 < 0.005) (*SI Appendix*, Table S9). Habituation effects thus did not impact significantly the amplitude tracking results.

3.2.2. Phase tracking

To assess phase synchronization we followed the same procedure as for newborns. Mean phase-coherence values across the whole group of infants were 0.312 (SD: 0.0181; range across channels: 0.276−0.336) for English, 0.288 (SD: 0.0288; range across channels: 0.253−0.351) for French, and 0.300 (SD: 0.0295; range across channels: 0.243−0.342) for Spanish (see *SI Appendix*, Table S11). Our results are of a similar magnitude to those observed in newborns above, and in adults (Pefkou et al., 2017).

To establish phase tracking, we assessed if the phase-coherence distributions were significantly different from zero (one-sample t-tests, two-tailed), and from chance (permutation test) (*SI Appendix*, Table S11, Table S12). We conducted a permutation analysis, similar to the one implemented for the newborn data. Fig. 4b indicates the channels where phase-coherence was significantly higher than zero and chance. The phase tracking results from the measured data were significant (higher than zero and the chance permutation results) for every channel and language.

To assess possible differences across languages and channels, we submitted the phase-coherence results from the 10 channels to a repeated measures ANOVA with *Language* and *Channel* as within-subjects factors. This ANOVA yielded no significant main effects of *Language* [F(2,696) = 1.579, p = 0.207] or *Channel* [F(9,696) = 1.474, p = 0.154], implying that the brain activity of 6 month-olds tracks the phase of the speech envelope equally well in the three languages in all channels.

4. Discussion

Here we have investigated the developmental trajectory of envelope tracking during the first months of life in order to better understand what role this mechanism plays in speech perception and language comprehension. We have shown that newborns, exclusively exposed to French prenatally, possess the neural capacity to track the amplitude and the phase of the speech envelope in their native language, as well as in rhythmically similar and different unfamiliar languages (Spanish and English). These findings reveal that envelope tracking represents a basic auditory ability that does not require extensive experience with speech, or knowledge of a given language (i.e. its grammar, or lexicon). They

may not even be specific to speech at all. These findings, therefore, support the hypothesis that speech envelope tracking is not sufficient for speech comprehension, although it may be a necessary prerequisite.

This is not to say, however, that envelope tracking is immune to developmental change. Indeed, we have found that at 6 months, infants exhibit a change in their ability to track the amplitude and the phase of the speech envelope. They no longer track the amplitude of the speech envelope in their native language French, but they keep up amplitude tracking for the unfamiliar languages, in particular for the rhythmically different language, English. The developmental trajectory of phase tracking differs in important ways from that of amplitude tracking: 6-month-olds continue to track the phase of the speech envelope in all three languages.

We hypothesize that the observed developmental changes from birth to 6 months likely reflect language development and may highlight different roles for amplitude and phase tracking. In particular, we suggest that the nature of envelope tracking/processing changes, from acoustic to linguistic, between birth and 6 months of age for the native language and to some extent the rhythmically similar unfamiliar language, whereas it remains acoustic for the rhythmically different unfamiliar language. Our hypothesis is in line with previous findings (Ortiz-Mantilla et al., 2013) showing theta activation when 6-month-olds were presented with non-native syllabic contrasts, compared to theta and gamma activation when faced with native syllabic contrasts.

The few existing studies on the development of amplitude tracking, taken together with our results suggest that the ability to track the amplitude of the speech envelope in the native language undergoes a U-shape trajectory during development: it is present at birth (as shown by our results), it is then absent around 6–7 months (as per our results and Kalashnikova et al., 2018), and it reappears later in development, as it has been found in older children (Abrams et al., 2008) and adults (Kubanek et al., 2013). Such U-shaped developmental trajectories are common in language acquisition (Gervain and Werker, 2008; Werker and Hensch, 2015) and are believed to result from perceptual and neural re-organization corresponding to attunement to the native language as processing native, but not unfamiliar, contrasts changes from acoustic to linguistic. In the light of this, we hypothesize that the absence of amplitude tracking to the native language that we observe at 6 months could reflect such perceptual re-organization, specifically a shift in attention from the syllabic units (present at 4−5 Hz and embedded in the envelope), known to be crucial for speech perception at birth (Bertoncini and Mehler, 1981; Bertoncini et al., 1995), towards the processing of phonemic units (present around 30 Hz), which become relevant for word and grammar learning, which begin around 6 months. More specifically, it has been suggested that starting at around 4 months infants start to rely more on the phoneme repertoire of their native language to discriminate languages within the same rhythmic class (Bosch and Sebastián-Gallés, 1997; Molnar et al., 2013), and starting at 6 months, infants start learning the first word forms (Tincoff and Jusczyk, 1999; Bergelson and Swingley, 2012) and grammatical structures (Gervain et al., 2008) of their native language, which require phoneme-level representations. We suggest that as a consequence of this, amplitude tracking subsides in the native language, as infants shift their focus from larger units (syllables) to smaller units (phonemes) in order to attune to the sound patterns of their native language. A prediction of this claim is that gamma oscillations, believed to be responsible for (sub)phonemic processing, are enhanced between birth and 6 months, especially for the native language – a prediction that is currently being investigated in our laboratory following upon existing work on this question (Peña et al., 2010; Ortiz-Mantilla et al., 2013; Nacar Garcia et al., 2018). Subsequently, amplitude tracking in the native language reappears, possibly once word and grammar learning are well underway. This empirical prediction may be tested by future research investigating amplitude tracking in 12-, 18- or 24-month-old infants.

Importantly, we found that 6-month-olds can track the amplitude of the speech envelope in unfamiliar languages (English and Spanish), a

condition that has never been tested before. This finding is in line with the above proposed account. For unfamiliar languages, the representational shift from the syllabic to the phonemic level does not occur, and processing remains acoustic, allowing amplitude tracking to take place. Relatedly, unfamiliar languages may attract greater attention from the participants (as suggested by Kalashnikova et al., 2018 for the infant-directed vs. adult-directed speech difference), which could have aided enhancing their amplitude tracking.

Phase tracking appears to follow a different developmental trajectory. It is present at birth as well as at 6 months, and into adulthood (Pefkou et al., 2017). Could this mean that amplitude and phase tracking represent two distinct neurophysiological mechanisms of oscillatory synchronisation (Meyer, 2018)? While both mechanisms track features of the speech envelope, it is possible that they play different roles in speech processing. Amplitude tracking (Nourski et al., 2009) as well as phase tracking (Luo and Poeppel, 2007; Peelle et al., 2013) have been related to speech comprehension, with enhanced envelope tracking for intelligible varieties of speech. Some studies however, have questioned this tracking–comprehension relationship by showing similar phase tracking responses for intelligible and unintelligible speech (time-reversed speech, Howard and Poeppel, 2010; Zoefel and VanRullen, 2016; time-compressed speech, Pefkou et al., 2017). These results thus call into question whether speech comprehension and envelope tracking are causally linked. Our newborn results showing amplitude and phase tracking of familiar and unfamiliar languages support the hypothesis that both mechanisms are present in the absence of comprehension. The finding that 6-month-olds track the phase of the speech envelope for familiar and unfamiliar languages equally well further shows that phase tracking is less modulated by language experience and comprehension than amplitude tracking. This converges well with adult studies showing phase tracking for unintelligible speech (Howard and Poeppel, 2010; Zoefel and VanRullen, 2016; Pefkou et al., 2017). Furthermore, phase and amplitude tracking of the envelope are observed in different frequency ranges: phase tracking mainly in the lower frequencies (delta and theta bands) (Luo and Poeppel, 2007; Howard and Poeppel, 2010; Peelle et al., 2013; Golumbic et al., 2013), while amplitude tracking mostly in the higher frequencies (usually including the gamma band) (Abrams et al., 2008; Nourski et al., 2009; Kubanek et al., 2013).

In the light of these results, we suggest that amplitude and phase tracking play different roles in speech processing. We suggest that phase tracking (present in the low frequencies) precedes and modulates amplitude tracking (present in the higher frequencies), by means of the nesting relationship that exists between low-frequencies and high-frequencies (i.e. the phase of the low frequencies modulates the amplitude of the high frequencies, Giraud and Poeppel, 2012). We speculate that phase tracking could represent an early, low-level, acoustic mechanism of speech processing, which in turn would explain why it takes place for a wider range of auditory stimuli (e.g. intelligible and unintelligible speech). Phase tracking could contribute to determine the feasibility of decoding the speech signal (intelligibility) and could drive amplitude tracking, which processes syllabic level, i.e. linguistic information, whenever relevant (when speech is familiar, forward-going, uncompressed etc.).

Note that the observed developmental differences are not likely to be due to the different states of alertness between the two groups (newborns were in quiet rest or asleep during the test, while 6-month-olds were awake) for at least two reasons. First, it has been shown that the auditory cortex is active during sleep and responds to sound (e.g. Sambeth et al., 2008). Sleeping newborns are even able to learn from auditory stimuli (Fifer et al., 2010). Second, differences in sleep states should impact all languages uniformly, whereas we have observed language-specific differences. Therefore, our developmental results are unlikely to be solely attributable to state of alertness differences.

Our experimental design involved stimulus repetitions, likely triggering habituation in infants' neural responses. However, our envelope tracking analyses rely on the general shape (amplitude tracking) and

phase (phase tracking) of the neural responses, with magnitude decreases not impacting strongly our measures.

5. Conclusion

Our study demonstrates that speech envelope tracking, a mechanism argued to play a foundational role in speech processing, is present from birth, and it takes place in the absence of both attention and comprehension (linguistic knowledge). Envelope tracking may thus underlie infants' sophisticated speech perception abilities, and may provide one of the mechanisms affording early brain plasticity for speech perception and language acquisition (Werker and Hensch, 2015). Our findings also suggest that amplitude and phase tracking are different neural mechanisms, impacted differently by linguistic experience and playing different roles in speech perception.

Data availability

The data that support the findings of this study are available from the corresponding author upon request.

Code availability

The custom codes that support the findings of this study are available from the corresponding author upon request.

Author contributions

M.O. and J.G. conceived and designed the experiment. M.O. performed the experiment. M.O. and R.G analysed the data. M.O., R.G. and J.G. wrote the paper. All the authors evaluated the results and edited the manuscript.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at https://doi.org/10.1016/j.dcn.2021.100915.

References

Abboub, N., Nazzi, T., Gervain, J., 2016. Prosodic grouping at birth. Brain Lang. 162, 46–59.
Abrams, D.A., Nicol, T., Zecker, S., Kraus, N., 2008. Right-hemisphere auditory cortex is dominant for coding syllable patterns in speech. J. Neurosci. 28, 3958–3965.
Ahissar, E., Nagarajan, S., Ahissar, M., Protopapas, A., Mahncke, H., Merzenich, M.M., 2001. Speech comprehension is correlated with temporal response patterns recorded from auditory cortex. Proc. Natl. Acad. Sci. U. S. A. 98, 13367–13372.
Benavides-Varela, S., Gervain, J., 2017. Learning word order at birth: a NIRS study. Dev. Cogn. Neurosci. 25, 198–208.

Benjamini, Y., Hochberg, Y., 1995. Controlling the false discovery rate: a practical and powerful approach to multiple testing. J. R. Stat. Soc. Ser. B Methodolog. 57 (1), 289–300.

Bergelson, E., Swingley, D., 2012. At 6–9 months, human infants know the meanings of many common nouns. Proc. Natl. Acad. Sci. 109, 3253–3258.

Bertoncini, J., Mehler, J., 1981. Syllables as units in infant speech perception. Infant Behav. Dev. 4, 247–260.

Bertoncini, J., Floccia, C., Nazzi, T., Mehler, J., 1995. Morae and syllables: rhythmical basis of speech representations in neonates. Lang. Speech 38 (4), 311–329.

Bosch, L., Sebastián-Gallés, N., 1997. Native-language recognition abilities in 4-month-old infants from monolingual and bilingual environments. Cognition 65, 33–69.

Ding, N., Patel, A.D., Chen, L., Butler, H., Luo, C., Poeppel, D., 2017. Temporal modulations in speech and music. Neurosci. Biobehav. Rev. 81, 181–187.

Drullman, R., Festen, J.M., Plomp, R., 1994a. Effect of temporal envelope smearing on speech reception. J. Acoust. Soc. Am. 95, 1053–1064.

Drullman, R., Festen, J.M., Plomp, R., 1994b. Effect of reducing slow temporal modulations on speech reception. J. Acoust. Soc. Am. 95, 2670–2680.

Fifer, W.P., Byrd, D.L., Kaku, M., Eigsti, I.M., Isler, J.R., Grose-Fifer, J., Tarullo, A.R., Balsam, P.D., 2010. Newborn infants learn during sleep. Proc. Natl. Acad. Sci. 107, 10320–10323.

Gerhardt, K.J., Abrams, R.M., Oliver, C.C., 1990. Sound environment of the fetal sheep. Am. J. Obstet. Gynecol. 162, 282–287.

Gervain, J., Werker, J.F., 2008. How infant speech perception contributes to language acquisition. Lang. Linguist. Compass 2 (6), 1149–1170.

Gervain, J., Werker, J.F., 2013. Prosody cues word order in 7-month-old bilingual infants. Nat. Commun. 4, 1–6.

Gervain, J., Nespor, M., Mazuka, R., Horie, R., Mehler, J., 2008. Bootstrapping word order in prelexical infants: a Japanese–Italian cross-linguistic study. Cogn. Psychol. 57, 56–74.

Giraud, A.L., Poeppel, D., 2012. Cortical oscillations and speech processing: emerging computational principles and operations. Nat. Neurosci. 15, 511.

Golumbic, E.M.Z., Ding, N., Bickel, S., Lakatos, P., Schevon, C.A., McKhann, G.M., Goodman, R.R., Emerson, R., Mehta, A.D., Simon, J.Z., Poeppel, D., 2013. Mechanisms underlying selective neuronal tracking of attended speech at a "cocktail party". Neuron 77, 980–991.

Hardware: actiCAP and actiCHamp, (32 channels) [Apparatus]. (2019), Gilching, Germany. Brain Products GmbH.

Howard, M.F., Poeppel, D., 2010. Discrimination of speech stimuli based on neuronal response phase patterns depends on acoustics but not comprehension. J. Neurophysiol. 104, 2500–2511.

Hurtado, J.M., Rubchinsky, L.L., Sigvardt, K.A., 2004. Statistical method for detection of phase-locking episodes in neural oscillations. J. Neurophysiol. 91, 1883–1898.

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. Sci. Rep. 8, 13745.

Kubanek, J., Brunner, P., Gunduz, A., Poeppel, D., Schalk, G., 2013. The tracking of speech envelope in the human cortex. PLoS One 8, e53398.

Kuhl, P.K., Williams, K.A., Lacerda, F., Stevens, K.N., Lindblom, B., 1992. Linguistic experience alters phonetic perception in infants by 6 months of age. Science 255, 606–608.

Lecanuet, J.P., Schaal, B., 2002. Sensory performances in the human foetus: a brief summary of research. Intellectica 34, 29–56.

Luo, H., Poeppel, D., 2007. Phase patterns of neuronal responses reliably discriminate speech in human auditory cortex. Neuron 54, 1001–1010.

Mampe, B., Friederici, A.D., Christophe, A., Wermke, K., 2009. Newborns' cry melody is shaped by their native language. Curr. Biol. 19 (23), 1994–1997.

McMurray, B., Aslin, R.N., 2005. Infants are sensitive to within-category variation in speech perception. Cognition 95 (2), B15–B26.

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

Mehler, J., Dupoux, E., Nazzi, T., Dehaene-Lambertz, G., 1996. Coping with linguistic diversity: the infant's viewpoint. Signal to Syntax: Bootstrapping From Speech to Grammar in Early Acquisition, pp. 101–116.

Meyer, L., 2018. The neural oscillations of speech processing and language comprehension: state of the art and emerging mechanisms. Eur. J. Neurosci. 48, 2609–2621.

Molnar, M., Gervain, J., Carreiras, M., 2013. Within-rhythm class native language discrimination abilities of Basque-Spanish monolingual and bilingual infants at 3.5 months of age. Infancy 19, 326–337.

Moon, C., Cooper, R.P., Fifer, W.P., 1993. Two-day-olds prefer their native language. Infant Behav. Dev. 16, 495–500.

Nacar Garcia, L., Guerrero-Mosquera, C., Colomer, M., Sebastian-Galles, N., 2018. Evoked and oscillatory EEG activity differentiates language discrimination in young monolingual and bilingual infants. Sci. Rep. 8 (1), 2770.

Nazzi, T., Bertoncini, J., Mehler, J., 1998. Language discrimination by newborns: toward an understanding of the role of rhythm. J. Exp. Psychol. Hum. Percept. Perform. 24, 756–766.

Nourski, K.V., Reale, R.A., Oya, H., Kawasaki, H., Kovach, C.K., Chen, H., Howard, M.A., Brugge, J.F., 2009. Temporal envelope of time-compressed speech represented in the human auditory cortex. J. Neurosci. 29, 15564–15574.

Ortiz-Mantilla, S., Hämäläinen, J.A., Musacchia, G., Benasich, A.A., 2013. Enhancement of gamma oscillations indicates preferential processing of native over foreign phonemic contrasts in infants. J. Neurosci. 33 (48), 18746–18754.

Peelle, J.E., Gross, J., Davis, M.H., 2013. Phase-locked responses to speech in human auditory cortex are enhanced during comprehension. Cereb. Cortex 23, 1378–1387.

Pefkou, M., Arnal, L.H., Fontolan, L., Giraud, A.L., 2017. θ-Band and β-band neural activity reflects independent syllable tracking and comprehension of time-compressed speech. J. Neurosci. 37, 7930–7938.

Peña, M., Pittaluga, E., Mehler, J., 2010. Language acquisition in premature and full-term infants. Proc. Natl. Acad. Sci. 107 (8), 3823–3828.

Ramus, F., Hauser, M.D., Miller, C., Morris, D., Mehler, J., 2000. Language discrimination by human newborns and by cotton-top tamarin monkeys. Science 288, 349–351.

Rosen, S., 1992. Temporal information in speech: acoustic, auditory and linguistic aspects. Philos. Trans. R. Soc. Lond. Series B Biol. Sci. 336 (1278), 367–373.

Sambeth, A., Ruohio, K., Alku, P., Fellman, V., Huotilainen, M., 2008. Sleeping newborns extract prosody from continuous speech. Clin. Neurophysiol. 119, 332–341.

Shannon, R.V., Zeng, F.G., Kamath, V., Wygonski, J., Ekelid, M., 1995. Speech recognition with primarily temporal cues. Science 270, 303–304.

Song, J., Iverson, P., 2018. Listening effort during speech perception enhances auditory and lexical processing for non-native listeners and accents. Cognition 179, 163–170.

Stefanics, G., Háden, G.P., Sziller, I., Balázs, L., Beke, A., Winkler, I., 2009. Newborn infants process pitch intervals. Clin. Neurophysiol. 120, 304–308.

Sun, J., Hong, X., Tong, S., 2011. A comparison of surrogate tests for phase synchronization analysis of neural signals. In: Proc. Asia-Pacific Signal Inf. Process. Assoc. Annu. Summit Conf, Vol. 203.

Tincoff, R., Jusczyk, P.W., 1999. Some beginnings of word comprehension in 6-month-olds. Psychol. Sci. 10, 172–175.

Tóth, B., Urbán, G., Háden, G.P., Márk, M., Török, M., Stam, C.J., Winkler, I., 2017. Large-scale network organization of EEG functional connectivity in newborn infants. Hum. Brain Mapp. 38, 4019–4033.

Tsuji, S., Cristia, A., 2014. Perceptual attunement in vowels: a meta-analysis. Dev. Psychobiol. 56, 179–191.

Varnet, L., Ortiz-Barajas, M.C., Erra, R.G., Gervain, J., Lorenzi, C., 2017. A cross-linguistic study of speech modulation spectra. J. Acoust. Soc. Am. 142, 1976–1989.

Vince, M.A., Armitage, S.E., 1980. Sound stimulation available to the sheep foetus. Reprod. Nutr. Développement 20, 801–806.

Werker, J.F., Hensch, T.K., 2015. Critical periods in speech perception: new directions. Annu. Rev. Psychol. 66, 173–196.

Werker, J.F., Tees, R.C., 1984. Cross-language speech perception: evidence for perceptual reorganization during the first year of life. Infant Behav. Dev. 7 (1), 49–63.

Zoefel, B., VanRullen, R., 2016. EEG oscillations entrain their phase to high-level features of speech sound. NeuroImage 124, 16–23.

Zoefel, B., Archer-Boyd, A., David, M.H., 2018. Phase entrainment of brain oscillations causally modulates neural responses to intelligible speech. Curr. Biol. 28 (3), 401–408 e5.