Explicit access to phonetic representations in 3-month-old infants

Karima Mersad, Claire Kabdebon, Ghislaine Dehaene-Lambertz

Mersad, K., Kabdebon, C., & Dehaene-Lambertz, G. (2021). Explicit access to phonetic representations in 3-month-old infants. Cognition, 104613. doi:10.1016/j.cognition.2021.104613

Explicit access to phonetic representations in 3-month-old infants

Karima Mersad *, Claire Kabdebon, Ghislaine Dehaene-Lambertz

Cognitive Neuroimaging Unit, CNRS, INSERM, CEA, Université Paris-Saclay, NeuroSpin Center, 91191 Gif/Yvette, France

A B S T R A C T

Preverbal infants are particularly good at discriminating syllables that differ by a single phoneme but do they perceive syllables as a whole unit or can they become aware of the underlying phonemes if their attention is attracted to the relevant level of analysis? We trained 3-month-old infants to pair two consonants, co-articulated with different vowels, with two visual shapes. Using event-related potentials, we showed that infants generalize the learned associations to new syllables with respect to the training phase. The systematic pairing of a visual label with a phonetic category is rapidly learned in a few trials, suggesting that phonemes are natural categories for infants but also that phonetic representations are accessible to internal operations outside the linguistic system. Hence, the possibility of an explicit access to the phonetic level, which is the main process underlying alphabetic reading system, is grounded in the early faculties of the human infant.

1. Introduction

Phonemes are at the base of the combinatorial power of human language, offering endless possibilities of meaningful sequences to communicate facts and ideas. If, on the speech production side, they can be relatively well described through a combination of binary features corresponding to idealized articulatory movements, their acoustic variability, and thus their perceptual relevance has long been the subject of debates in which Jacques Mehler was one of the main protagonists. His originality was to draw argument from preverbal infants' abilities. For him, infants were a key to understand the nature of human language because their abilities could not be explained by environmental tuning, especially in the case of newborns who are just exposed to aerial speech, nor by a motor component, given their weak and immature articulatory skills. Thus, initial speech representations have been the subject of many of his publications.

While many authors (Kuhl, 2004; Werker & Tees, 2005) agree that phonetic representations are computed by infants at the end of their first year because all properties defining phonetic perception in adults are observed (i.e. categorisation, normalisation through irrelevant acoustic cues, such as voices, emotions and co-articulation, native-language dependence), the status of phoneme perception in the early months remained underspecified (Hallé & Cristia, 2012; Mehler, Dommergues, Frauenfelder, & Segui, 1981; Räsänen, Doyle, & Frank, 2018). The debate is whether infants initially perform only general spectro-temporal analyses centred on the syllable nucleus, vowels and consonants being blended in an underspecified sequence, or whether they can already compute discrete phonetic representations. For infants to discover the phonology, words and rules of the native language, the refinement of the initial representations of speech is not without consequence, multiplying the size of the set to be analysed from a relatively small set of phonemes to a much larger set of syllables or even to multiple underspecified spectro-temporal patterns.

Since co-articulation strongly affects the spectro-temporal pattern obliterating any simple correlation between acoustic cues and phonetic categories, the observation that infants are able to recognize the same consonant even when coarticulated with different vowels is one of the arguments presented in favour of independent phonetic representations beyond the syllable. Using a non-nutritive sucking procedure, Mehler and colleagues habituated 2-month-old infants to a set of consonant-vowel (CV) syllables sharing the same consonant but with different vowels, then introduced a new CV token, in which either the vowel, the consonant or both were new (Bertoncini, Bijeljac-Babic, Jusczyk, Kennedy, & Mehler, 1988; Jusczyk & Derrah, 1987). The two-month-olds reacted to all changes. Using pupillometry, Hochmann and Papeo (2014) reported that 6-month-olds, but not 3 month-olds, noticed that the same consonant was shared in a set of CV syllables. Using event-related potentials (ERPs), we recorded a mismatch response in 3-month-olds when there was a change of the first consonant in short series of CV syllables despite the fact that each syllable comprised a different vowel (Mersad & Dehaene-Lambertz, 2015). Finally, using multivariate decoding on ERPs in 3 month-old infants listening to 120

* Corresponding author.
E-mail address: Karima.mersad@parisdescartes.fr (K. Mersad).

different CV syllables sharing either the place or the manner of articulation, Gennari et al. (submitted) reported that the syllables were first encoded along the orthogonal dimensions of the phonetic features (place and manner in this case), which were subsequently aggregated in a second stage, corresponding to the consonant of the CV syllable. This result was congruent with adults' electro-corticographic (ECOG) results showing that some electrodes over the superior temporal gyrus were exquisitely sensitive to phonetic features such as manner of articulation, sonority, etc. (Mesgarani, Cheung, Johnson, & Chang, 2014).

While multivariate analyses reveal that in infants, as in adults, speech sounds elicit distributed patterns of brain activity which can be distinguished along relevant linguistic features, this distributed information may remain encapsulated, implicitly encoded across large neuronal populations without the possibility of an explicit readout of the phonetic representation for further cognitive processes. For example, the voice onset time dimension might already be decodable from the brainstem, since the neural response is faithfully following the auditory input (Johnson et al., 2008). However, it does not mean that at this early stage of the processing hierarchy, a phonetic representation is available to the brain. Similarly, sucking recovery, changes in pupil diameter and ERP mismatch responses after a change of phoneme in a set of syllables might only reflect the distance between the new syllable and the previous set (Garrido, Sahani, & Dolan, 2013), rather than signaling a genuine representation of a categorical difference. It is indeed possible that in a multi-dimensional acoustic-phonetic space, syllables that share more features might be perceptually closer than those that share fewer features (Pierrehumbert, 2003).

Therefore, how can we assess whether pre-verbal infants have access to phonetic representations? The operative difference between an implicit encapsulated neural pattern and an explicit mental representation is the availability of the representation for further cognitive operations (Bouton et al., 2018). The distributed neural pattern is then summarized in an explicit neural code that can be shared, an operation called by Karmiloff-Smith (1986) a representational redescription (note that in this framework, "explicit" does not mean "conscious" but "manipulable"). This code can be ultimately translated in another modality, e.g. phoneme to grapheme as it is done in alphabetic reading. Thus, alphabetic reading systems provide the ultimate proof that the speech input can be readout as a stable succession of phonemes to which visual shapes are robustly and consistently associated.

To find out whether infants are able to access and manipulate phonemes within a syllable, and thus have access to explicit phonetic representations beyond implicit neural patterns, we designed a new ERP task based on the underlying principle of alphabetic reading (phoneme-visual shape association). Three-month-old infants were trained to associate a syllable /bX/ (or /gX/) with a vowel X randomly drawn from the set (/i/, /ɛ/, /ã/, /ɛ̃/), with a yellow "+" image (or a red "%"). In a subsequent test phase, we examined whether infants generalize the association to two novel syllables /ba/ and /ga/, which were associated with both shapes with equal probability (Fig. 1). Three types of regularities can be extracted from this audiovisual training. Firstly, each of the 8 syllable-shape pairs could be memorized separately. This is however well beyond infants' abilities at this age (Gogate, 2010; Gogate, Prince, & Matatyaho, 2009). In addition, with this information, the new syllables /ba/ and /ga/ cannot be associated with either of the two shapes. Secondly, infants could use the global similarity of the spectro-temporal patterns between syllables to create two sets to which each of the visual shapes is attached. However, the spectro-temporal pattern of place of articulation is variable depending on the co-articulated vowel, and it is notably different between /a/ and the vowels of the training set. Lastly, infants could learn the association between the visual shape and the common consonant. In this case, learning is limited to two pairs and infants should exhibit a mismatch response for an incongruent pairing, even in the context of a novel vowel. Note that in this paradigm, unlike the previous tasks using discrimination, we test a second order level. The measure pertains to the congruity of the visual shape with the consonant, which implies a stable phonetic representation that can be shared outside the linguistic system.

2. Materials and method

2.1. Participants

Thirty-three 3-month-old full-term infants raised in a French speaking environment were tested, but 5 did not complete the familiarization phase and 8 did not provide exploitable data due to poor data quality (see analysis section for exclusion criteria). The remaining 20 subjects comprised 8 girls and 12 boys, with a mean age of 14 weeks and 3 days (13 w. 5 d. to 16 w. 4 d.). The study was approved by the regional ethical committee for biomedical research, and the parents gave their written informed consent for the protocol.

2.2. Auditory stimuli

Ten syllables were naturally produced by a French female speaker (/gɛ/: 235 ms /, /gã/: 298 ms, /gɛ̃/: 272 ms, /gi/: 213 ms, /bɛ/: 244 ms, /bã/: 244 ms, and /bɛ̃/: 245 ms, /bi/: 214 ms). We kept their initial duration to increase the variability of stimuli during familiarization. Two other syllables (/ba/: 199 ms and /ga/: 194 ms) produced by the same speaker and matched in duration were used as test stimuli. All syllables were matched for subjective intensity. Note that the vowels {/i/, /ɛ/, /ã/, /ɛ̃/} were chosen on the one hand to increase the variability of the spectro-temporal patterns of the consonant due to co-articulation with a place of articulation feature, and on the other hand to present a spectro-temporal pattern different from /a/. At a phonetic level, the closest vowel to /a/ in the training set is /ã/. Trehub (1976) reported that 5–17 week-olds infants discriminate the contrast /pa/ −/pã/ (see also Bosch & Sebastián-Gallés, 2003; Kuhl, 1983; Swoboda,

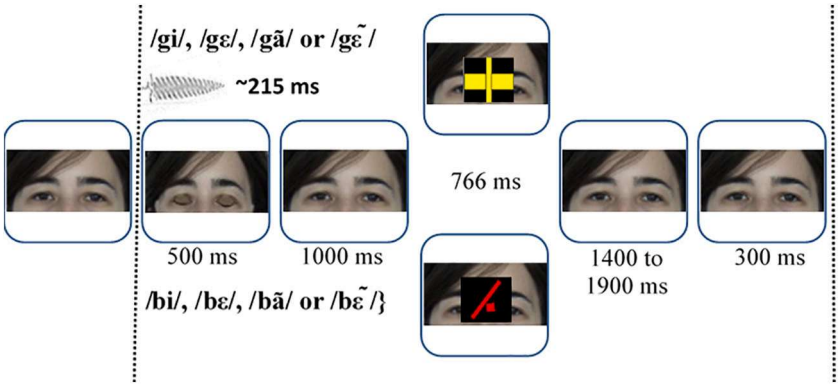

Fig. 1. Trial structure. A syllable /bX/, or /gX/, with X randomly chosen from {/i/, /ɛ/, /ã/, /ɛ̃/} is synchronously presented with blinking eyes, followed by a geometrical shape presented over the eyes 1500 ms after the syllable onset. One consonant is associated with one shape (e.g., /g/ with the yellow "+" and /b/ with the red "%"), counterbalanced across infants. During the test phase, two new syllables /ba/ and /ga/ are added to the set, associated with either of the two shapes with a fifty-fifty probability. The correct vs. incorrect association between the shape and the consonant defines congruent vs. incongruent trials. Incongruency can only be perceived if infants are able to notice the constant association between a consonant and a shape despite vowel variation. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

Kass, Morse, & Leavitt, 1978; Trehub, 1973 for other examples of discrimination for more or less close vowels in infants). Furthermore, all vowels are present in French and thus should be perceived as different by our 14-week-old French infants even if some perceptual narrowing toward the native language might already be present (Moon, Lagercrantz, & Kuhl, 2013).

Stimuli were recorded on the left channel and a click was positioned on the right channel at the exact time-point of syllable onset to be used as stimulus onset (TTL) for the EEG recordings. This channel was connected to EEG amplifiers through the DIN port to create a TTL signal whereas the left channel was connected to an audio amplifier to present the syllable through a speaker facing the infant. Hence, when the sound was played by the PC soundcard, the brain voltage and the trigger signal were recorded simultaneously with the same temporal resolution.

2.3. Visual Stimuli

The shapes to learn were a yellow cross and a red square tangent to a red rectangular bar. To keep infant's attention toward the screen, we also used two still images of a woman, one with closed eyes and the second with open eyes. Only the upper part of the face (hair, forehead, and eyes) was presented against a blue background.

2.4. Experimental design

To give infants a direction of the trials time-line (Fig. 1), trials always began with an eye blink: The picture with closed eyes was presented during 500 ms while a syllable randomly chosenpl from a set of 4 syllables was presented auditively, followed by the open eyes picture for one second. Then a shape appeared in the middle of the eyes during 766 ms. The open eyes were then visible again until the next trial onset marked by the reappearance of the closed eyes picture. This inter-trial period had a random duration (3200 to 3700 ms). The jitter in inter-trial duration was used to increase the perception of a sound-image pair, and avoid wrong associations between a shape and the following syllable. The eyes had two functions in the paradigm, firstly to act as a fixation point and avoid eye movements and secondly to act as a communication cue to encourage infants in the task (Csibra & Gergely, 2009). In our previous study in which infants were discriminating syllables while a face was either blinking or making an articulatory movement, the mismatch response was paradoxically larger for the blink than for the articulatory situation suggesting that the blink was acting as an attentional cue (Mersad & Dehaene-Lambertz, 2015).

During familiarization, infants were first trained for 10 trials to one pairing (e.g. syllable /bX/ and the yellow "+"), then to the other pairing for 10 trials (e.g. syllable /gX/ followed by a red "%"), then both pairings were presented 5 times each in a random order. The order of presentation (/b/ or /g/) and phoneme-shape associations were counterbalanced across infants.

After the 30 trials of familiarization, the test phase began and the new syllables /ba/ and /ga/ were added to the previous set of syllables. These two syllables were followed half of the time by one shape and the other half by the other shape. The test comprised 20 short blocks of 8 randomly presented trials. Each one comprised four familiar trials (2 /bX/ and 2 /gX/) to maintain learning and four new trials (2 /ba/ and 2 /ga/) followed with a shape either congruent or incongruent with learning. No two incongruent trials followed one another. We used short blocks to avoid imbalance in the number of trials between conditions over time. Thus, in total, test trials consisted of 80 always correct syllables-shape pairs (those from the training set) and 40/40 congruent/incongruent trials using the new syllables /ba/ and /ga/. Regarding the consonant, 75% of the trials were congruent and 25% were incongruent and should elicit surprise.

2.5. EEG recording and pre-processing

The 129-channel Geodesic sensor net (EGI) was placed on the infant head relative to anatomical marks with the infant seated on the parents' laps. Then, the infant and parent were seated facing a screen and a loudspeaker positioned behind the screen in a soundproof booth. The screen was located approximately 80 cm away from the infant's face. A video allowed the experimenter to manage the infants' attention outside the booth. If the infant looked away from the screen, the experiment was paused and the infant's gaze was guided back to the screen before the experiment resumed. If it was not possible, the experiment was terminated.

Scalp voltages were referenced to the vertex, amplified (EGI, amplifiers N200), and digitized at 250 Hz. After recording, the data were band-pass filtered [0.2-15 Hz], then segmented into epochs starting 500 ms before and ending 1800 ms after the onset of the shape presentation. Artifact detection was performed with custom-made scripts based on the EEGLAB toolbox (Delorme & Makeig, 2004). Epochs were considered unsuitable for analysis if their fast average amplitude exceeded 250 μV or their deviation between fast and slow running averages exceeded 150 μV. In each subject, channels that had more than 75% of epochs marked as bad were considered as bad channels. Five channels surrounding the eyes (17,125,126, 127 and 128) were rejected in several infants due to fixation difficulties. Therefore, we systematically rejected them in all subjects. For the remaining channels, 1.7 channels on average were rejected per infant. Trials having more than 50% of bad channels were rejected from the analysis. When a trial showed more than 25% of bad channels, these data were replaced by interpolating from neighbouring electrodes using spherical splines. After the artifact detection process, we obtained on average 81 acceptable trials per infant for the entire experiment (familiarization + test) with a mean number of rejected trials of 25.25. For the two conditions of interest (congruent/incongruent new syllables), we obtained 13.7 and 13.6 usable trials (5 to 25 trials). Eight infants were rejected because they had less than 5 artifact-free trials in one of the two conditions. The retained epochs were re-referenced at each data-point to the mean voltage (Mersad & Dehaene-Lambertz, 2015) to obtain a reference-free average and were finally baseline-corrected (−200 ms to −2 ms) in each subject.

2.6. Statistical analysis

The experimental design is based on a correct, or incorrect, association of a visual shape with the new syllables /ba/ and /ga/. We thus focused our analyses on the visual ERP, expecting a modulation of its amplitude in function of the congruency of the shapes with the infant's expectations, as reported by Kouider et al. (2015), Kabdebon and Dehaene-Lambertz (2019), Emberson, Richards, and Aslin (2015). Because infant's confidence in predicting the image depends on the level of learning and consolidation (Friedrich, Wilhelm, Mölle, Born, & Friederici, 2017; Kidd, Piantadosi, & Aslin, 2012), it was difficult to predict the direction of the effect. However, Kouider et al. (2015) and Kabdebon et al. (2019) reported larger responses for the expected stimulus.

We thus selected electrodes classically used to study visual ERPs: a cluster of 18 posterior electrodes (comprising TP9, PO7, O1, Oz, O2, PO8, TP10 and the electrodes underneath (Luu & Ferree, 2000)). We considered the responses evoked not only by the onset of the associated shape but also by the reappearance of the woman's eyes at the offset of the shape, because eyes robustly elicit evoked responses in infants (Farroni, Massaccesi, Pividori, & Johnson, 2004; Gliga & Dehaene-Lambertz, 2005). The time-windows were selected over the cluster of electrodes on the grand-average performed across all infants and all trials (Fig. 2). The visual N1, P1, N2, P400 are clearly visible after each of the two changes (first change (ev1): going from eyes to shape, second change (ev2): shape to eyes), superimposed on a slow wave with a signal returning to baseline only around 1.7 s after the shape onset. The peaks of the two largest components (P1 and P400) were at 124 ms and 464 ms

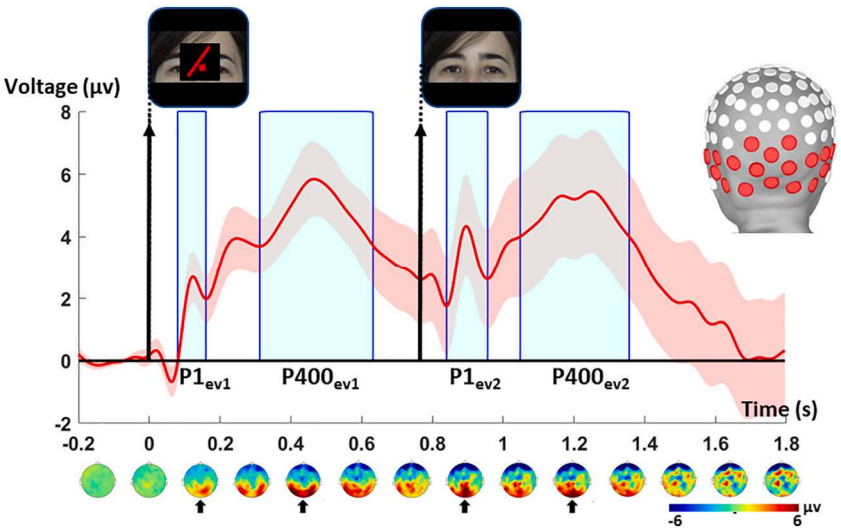

Fig. 2. Visual ERP to shape onset and offset. Grand-average of all trials of all conditions recorded from the channels identified in red on the net topography. P1 and P400 are visible after each picture, but superimposed on a slow wave developing during the whole epoch before the signal comes back to baseline-level around 1.7 s after the shape onset. Blue rectangles indicate the time-windows chosen to analyze the P1 and P400 evoked after each picture onset. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

after the shape onset and at 130 ms and 482 ms after the eyes onset.

In each infant and for each condition, we averaged the voltage over the cluster of electrodes and over the time-windows centred on these events (between two troughs for P1, and between the through and the same voltage value on the other side of the peak for P400, i.e. 80–160 ms, 312–632 ms post shape onset and 74–190 ms, 282–590 ms post-eyes onset (corresponding to 840–956 ms and 1.048–1.356 ms post shape onset). We entered these values in 4 different *t*-test analyses (one for each time-window) with congruent vs. incongruent factors as a within-subject factor (Bonferroni correction p_corr <0.0125).

3. Results

There was no significant difference for the components following the shape onset ($P1_{ev1}$: $t(19) = 1.13$, $p = 0.27$; $P400_{ev1}$: $t(19) = 1.61$, $p = 0.12$), contrary to those following the eyes reappearance ($P1_{ev2}$: $t(19) = 3.89$, $p = 0.001$; $P400_{ev2}$: $t(19) = 3.21$, $p = 0.005$, see Table 1). As can be seen in Fig. 3, the $P1_{ev2}$ and $P400_{ev2}$ were engulfed in a slow wave. Thus to better analyze a possible difference of amplitude in these components, we realigned the signal on the 100 ms preceding the onset of the eyes reappearance (baseline: 666–766 ms post shape onset), the amplitude remained different for the $P1_{ev2}$ ($t(19) = 2.53$, $p = 0.020$) but not the P400 ($t(19) = 1.18$, $p = 0.25$, Table 1, Fig. 3C).

4. Discussion

After a short training of 30 trials during which 3-month-old infants learned to associate two consonants presented in different CV syllables paired with two specific shapes, infants displayed significantly different visual ERPs for congruent and incongruent consonant-shape pairings. Importantly, the consonants were presented in novel syllables different from those presented during the learning phase, and were paired with either of the two shapes with equal probability. As a result, no association between the novel syllables and the shapes could have been learned during the test phase. Instead, to succeed, infants had to represent the abstract identity of the first consonant in order to discover the systematic association with the following visual shape.

As can be seen in Fig. 3, two effects superimposed over time. On the one hand, a difference between conditions that slowly increased from the time-window of the first P400 to almost the end of the epoch, due to a positive drift in congruent trials and a weaker negative drift in incongruent trials. On the other hand, a tendency to a weaker amplitude of the visual ERPs (P1, P400) in incongruent trials, notably when the eyes reappear on the screen. Slow waves signal attentional engagement after a surprising event (Csibra, Kushnerenko, & Grossmann, 2008; Dehaene-Lambertz & Dehaene, 1994) but also expectation of a subsequent event (Kouider et al., 2015; Valenza & Mento, 2016). Usually a direct gaze evokes large ERPs in infants (Farroni et al., 2004; Gliga & Dehaene-Lambertz, 2005). Thus, the positive drift and the larger P1 to eyes in congruent trials might suggest that infants were waiting for the reappearance of the eyes. Using near infra-red spectroscopy, Emberson et al. (2015) reported activation of the visual cortex as a result of infants' expectation, even when the visual stimulus was omitted. In incongruent trials, infants were distracted by the previous erroneous shape, preventing them to attend to the trial timeline in a similar way. An attentional bottleneck is consistent with many studies reporting that infants have difficulty perceiving new stimuli during stimulus orienting (Richards, 1997; Tsurumi, Kanazawa, Yamaguchi, & Kawahara, 2019).

In the present study, the congruency effect unfolded slowly. However, in some other audio-visual pairing tasks with older infants, an early priming effect, characterised by larger early ERP components for the expected visual item, has been reported (Kabdebon & Dehaene-Lambertz, 2019; Kouider et al., 2015). The delay observed here might be due to the difficulty of the task but also to the fact that infants were younger here than in these studies. The myelination of long-range cortico-cortical tracts, necessary for fast transfer of information between brain regions (Adibpour, Dubois, & Dehaene-Lambertz, 2017) and efficient priming effects on early processes through top-down connections is gradual during the first year of life.

Table 1
Mean, SE and Effect size of the $P1_{ev2}$ and $P400_{ev2}$ (bl = baseline).

Component	Condition	Mean (μv)	SE (μv)	Effect size (Cohen's d)
$P1_{ev2}$ (bl = [−200 0] ms)	Congruent	11.47	3.44	0.869
	Incongruent	−0.23	2.64	
$P400_{ev2}$ (bl = [−200 0] ms)	Congruent	13.18	4.1	0.718
	Incongruent	2.54	2.49	
$P1_{ev2}$ (bl = [666 766] ms)	Congruent	3.89	1.42	0.567
	Incongruent	−0.16	1.14	
$P400_{ev2}$ (bl = [666 766] ms)	Congruent	5.61	2.51	0.264
	Incongruent	2.62	2.12	

4.1. Access to phonetic representations in infants?

Because phonemes are the smallest units that can differentiate the meanings of words, it was hypothesized that infants begin to form phonetic representations once they notice that these small variations are relevant for distinguishing words (Best, 1993; MacKain, 1982; Werker &

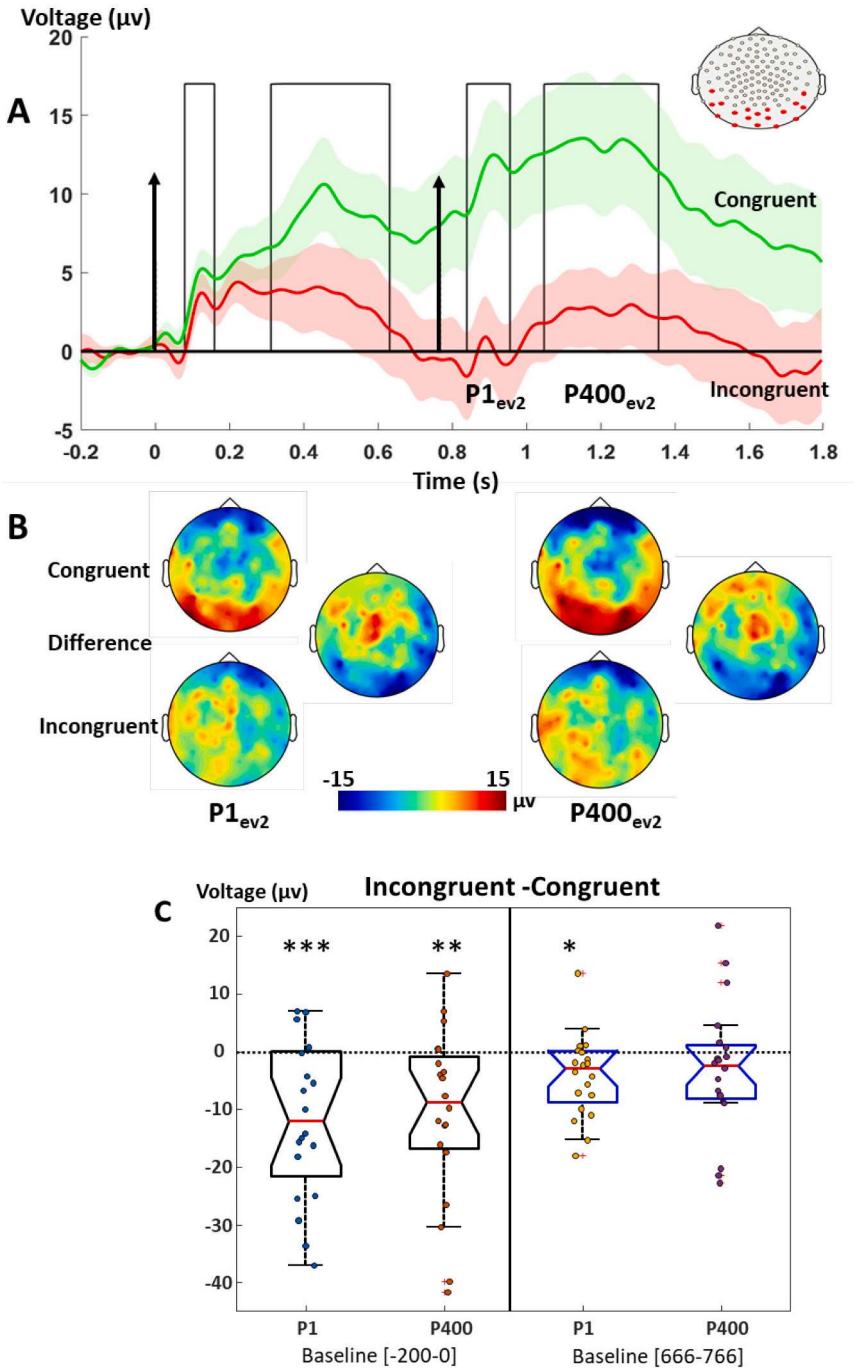

Fig. 3. ERPs in response to congruent versus incongruent shapes and their topographies. (A) Grand-average in the congruent (green curve) and incongruent (red curve) conditions recorded on the channels identified on the net topography. Rectangles indicate the time-windows of the analysed ERP components. (B) Topographical maps of the grand average in each condition and their difference, during the $P1_{ev2}$ and $P400_{ev2}$ time-windows (C) Scatterplot and boxplot (median, upper and lower quartiles. The whiskers indicate the 25th and the 75th percentiles) of the difference between conditions for each component. Each point represents one infant. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

Tees, 1984, 1999). Our study refutes this hypothesis and is congruent with other observations suggesting that phonemes have an earlier perceptual relevance. 6–8 month-olds adapt their phonetic discrimination responses to the statistics of their environment (Maye, Werker, & Gerken, 2002), and 9 month-olds use contrastive pairings between objects and sounds to discover new phonetic categories (Yeung & Werker, 2009). Finally, it has been suggested that allophones (i.e., different realizations of the same phoneme with no bearing on meaning) can be learned at a prelexical level (Dupoux & Peperkamp, 2002) due to their complementary statistical distributions in words. This model proposed by Dupoux & Peperkamp implies that phonemes are pertinent units whose distributions can be computed regardless of their position in words and the surrounding phonemes. An interesting follow-up of our study would be to investigate the effect of position in the syllable (onset vs. coda) and whether infants would recognize that /ba/ and /ab/ share the same phoneme /b/.

If our results show that infants have phonetic representations, they also go further by showing that these representations can be associated to arbitrary visual shapes. This successful associative learning indicates that beyond implicit distributed representations of phonetic features, the infant brain encoded consonant identity as an explicit mental variable available for further processes. As stated by Karmiloff-Smith, information already present in the system is re-described in a different representational format (Karmiloff-Smith, 1986, 1995). For example, learning to read leads to summarize multiple, distributed phonetic features under a unique visual symbolic representation. It should be noted that the conscious representation of speech as a sequence of phonetic units does not arise spontaneously during development but from specific

training in an alphabetic system (Morais, Cary, Alegria, & Bertelson, 1979). Even though illiterate adults can hear the difference between syllables, they are not aware of their elementary constituents and cannot easily decompose them into their phonetic segments nor easily exchange phonemes between syllables (e.g. convert John Lennon into Lohn Jennon). We propose here that the repeated presentation of visual items served as attentional cues for infants to notice that all syllables associated with the same shape shared a common attribute, namely the consonant. Similarly, assigning the same label to a set of dinosaur images help 3-month-old infants classify a new dinosaur image, but not a fish image, as belonging to the training set (Ferry, Hespos, & Waxman, 2013). In another study, 5-month-olds were able to use visual labels to categorize abstract structures (AAB words vs ABA words) (Kabdebon & Dehaene-Lambertz, 2019).

That three-month-olds can successfully map a set of phonetic exemplars onto the same visual label, well before they acquire their first words, reveals not only that phonemes are natural categories for preverbal infants but also that the same mechanism of directing attention toward the phonetic level can be elicited at this age as later on for reading acquisition. Along the same line, Lohvansuu, Hämäläinen, Ervast, Lyytinen, and Leppänen (2018) explored the relation between early phonetic representation and later reading abilities in children from dyslexic families. They reported that larger ERP amplitude to a frequently repeated speech stimulus at 6 months predicted faster reading speed at adolescence. The large ERP amplitudes to speech recorded early on, may reflect the efficient and stable encoding of explicit phonetic representations that later translates into proficient lexical access and phonological awareness.

Using decoding methods on infants' ERPs to CV syllables, Gennari et al. (submitted) reported two stages: a first one corresponding to the projection of the speech input on a space of phonetic features and a second stage where the projections are aggregated in a consonant-identity (i.e. phonetic level), which remained distinct from the vowel code. In the present study, we show that this second stage consonant representation is explicit. It remains to investigate the implicit vs. explicit nature of the first stage, for example by assigning a visual shape to a phonetic feature.

In conclusion, our study shows that preverbal infants who are only starting to vocalize can access sub-syllabic components and learn to map an initial consonant to an arbitrary visual shape in a few trials. Although our study was limited to two phonemes located at the onset of the syllable, this proof-of-concept reveals that the main cognitive process underlying reading, (i.e., creating a grapheme-phoneme pairing) is grounded in the natural faculties of the linguistic system.

Acknowledgements

Funding: This research was supported by the Fondation NrJ, Fondation de France, Fondation Bettencourt and the European Research Council (ERC) under the European Union 's Horizon 2020 research and innovation program (grant agreement No. 695710).

References

Adibpour, P., Dubois, J., & Dehaene-Lambertz, G. (2017). Right but not left hemispheric discrimination of faces in infancy. *Nature Human Behaviour, 1*. https://doi.org/10.1038/s41562-017-0249-4.

Bertoncini, J., Bijeljac-Babic, R., Jusczyk, P. W., Kennedy, L. J., & Mehler, J. (1988). An investigation of young infants' perceptual representations of speech sounds. *Journal of Experimental Psychology: General, 117*(1), 21.

Best, C. T. (1993). Emergence of language-specific constraints in perception of non-native speech : A window on early phonological development. In *Developmental Neurocognition : Speech and Face Processing in the First Year of Life* (pp. 289–304). Springer.

Bosch, L., & Sebastián-Gallés, N. (2003). Simultaneous bilingualism and the perception of a language-specific vowel contrast in the first year of life. *Language and Speech, 46* (2–3), 217–243. https://doi.org/10.1177/00238309030460020801.

Bouton, S., Chambon, V., Tyrand, R., Guggisberg, A. G., Seeck, M., Karkar, S., … Giraud, A.-L. (2018). Focal versus distributed temporal cortex activity for speech sound category assignment. *Proceedings of the National Academy of Sciences*, 201714279. https://doi.org/10.1073/pnas.1714279115.

Csibra, G., Kushnerenko, E., & Grossmann, T. (2008). Electrophysiological methods in studying infant cognitive development. In *Handbook of Developmental Cognitive Neuroscience* (2nd ed., pp. 247–262). MIT Press.

Dehaene-Lambertz, G., & Dehaene, S. (1994). Speed and cerebral correlates of syllable discrimination in infants. *Nature, 370*(6487), 292–295. https://doi.org/10.1038/370292a0.

Delorme, A., & Makeig, S. (2004). EEGLAB: An open source toolbox for analysis of single-trial EEG dynamics including independent component analysis. *Journal of Neuroscience Methods, 134*(1), 9–21.

Dupoux, E., & Peperkamp, S. (2002). Fossil markers of language development : Phonological 'deafnesses' in adult speech processing. *Phonetics, phonology, and cognition*, 168–190.

Emberson, L. L., Richards, J. E., & Aslin, R. N. (2015). Top-down modulation in the infant brain : Learning-induced expectations rapidly affect the sensory cortex at 6 months. *Proceedings of the National Academy of Sciences, 112*(31), 9585–9590. https://doi.org/10.1073/pnas.1510343112.

Farroni, T., Massaccesi, S., Pividori, D., & Johnson, M. H. (2004). Gaze following in newborns. *Infancy, 5*(1), 39–60.

Ferry, A. L., Hespos, S. J., & Waxman, S. R. (2013). Nonhuman primate vocalizations support categorization in very young human infants. *Proceedings of the National Academy of Sciences, 110*(38), 15231–15235. https://doi.org/10.1073/pnas.1221166110.

Friedrich, M., Wilhelm, I., Mölle, M., Born, J., & Friederici, A. D. (2017). The sleeping infant brain anticipates development. *Current Biology, 27*(15). https://doi.org/10.1016/j.cub.2017.06.070, 2374-2380.e3.

Garrido, M. I., Sahani, M., & Dolan, R. J. (2013). Outlier responses reflect sensitivity to statistical structure in the human brain. *PLoS Computational Biology, 9*(3), Article e1002999. https://doi.org/10.1371/journal.pcbi.1002999.

Gliga, T., & Dehaene-Lambertz, G. (2005). Structural encoding of body and face in human infants and adults. *Journal of Cognitive Neuroscience, 17*(8), 1328–1340.

Gogate, L. J. (2010). Learning of syllable–object relations by preverbal infants : The role of temporal synchrony and syllable distinctiveness. *Journal of Experimental Child Psychology, 105*(3), 178–197. https://doi.org/10.1016/j.jecp.2009.10.007.

Gogate, L. J., Prince, C. G., & Matatyaho, D. J. (2009). Two-month-old infants' sensitivity to changes in arbitrary syllable–object pairings : The role of temporal synchrony. *Journal of Experimental Psychology: Human Perception and Performance, 35*(2), 508–519. https://doi.org/10.1037/a0013623.

Hallé, P., & Cristia, A. (2012). Global and detailed speech representations in early language acquisition. In *Speech planning and dynamics* (pp. 11–38). Peter Lang. https://pure.mpg.de/pubman/faces/ViewItemOverviewPage.jsp?itemId=item_1381615.

Hochmann, J.-R., & Papeo, L. (2014). The invariance problem in infancy : A Pupillometry study. *Psychological Science, 25*(11), 2038–2046. https://doi.org/10.1177/0956797614547918.

Johnson, K. L., Nicol, T., Zecker, S. G., Bradlow, A. R., Skoe, E., & Kraus, N. (2008). Brainstem encoding of voiced consonant–vowel stop syllables. *Clinical Neurophysiology, 119*(11), 2623–2635. https://doi.org/10.1016/j.clinph.2008.07.277.

Jusczyk, P. W., & Derrah, C. (1987). Representation of speech sounds by young infants. *Developmental Psychology, 23*(5), 648.

Kabdebon, C., & Dehaene-Lambertz, G. (2019). Symbolic labeling in 5-month-old human infants. *Proceedings of the National Academy of Sciences, 116*(12), 5805–5810. https://doi.org/10.1073/pnas.1809144116.

Karmiloff-Smith, A. (1986). From meta-processes to conscious access : Evidence from children's metalinguistic and repair data. *Cognition, 23*(2), 95–147. https://doi.org/10.1016/0010-0277(86)90040-5.

Karmiloff-Smith, A. (1995). *Beyond modularity : A developmental perspective on cognitive science*. MIT Press.

Kidd, C., Piantadosi, S. T., & Aslin, R. N. (2012). The goldilocks effect : Human infants allocate attention to visual sequences that are neither too simple nor too complex. *PLoS One, 7*(5), Article e36399. https://doi.org/10.1371/journal.pone.0036399.

Kouider, S., Long, B., Le Stanc, L., Charron, S., Fievet, A.-C., Barbosa, L. S., & Gelskov, S. V. (2015). Neural dynamics of prediction and surprise in infants. *Nature Communications, 6*, 8537.

Kuhl, P. K. (1983). Perception of auditory equivalence classes for speech in early infancy. *Infant Behavior and Development, 6*(2), 263–285. https://doi.org/10.1016/S0163-6383(83)80036-8.

Kuhl, P. K. (2004). Early language acquisition : Cracking the speech code. *Nature Reviews Neuroscience, 5*(11), 831–843. https://doi.org/10.1038/nrn1533.

Lohvansuu, K., Hämäläinen, J. A., Ervast, L., Lyytinen, H., & Leppänen, P. H. T. (2018). Longitudinal interactions between brain and cognitive measures on reading development from 6 months to 14 years. *Neuropsychologia, 108*, 6–12. https://doi.org/10.1016/j.neuropsychologia.2017.11.018.

Luu, P., & Ferree, T. (2000). *Determination of the Geodesic Sensor Nets' Average Electrode Positions and Their 10–10 International Equivalents. Technical Note*.

MacKain, K. S. (1982). Assessing the role of experience on infants' speech discrimination. *Journal of Child Language, 9*(3), 527–542.

Maye, J., Werker, J. F., & Gerken, L. (2002). Infant sensitivity to distributional information can affect phonetic discrimination. *Cognition, 82*(3), B101–B111.

Mehler, J., Dommergues, J. Y., Frauenfelder, U., & Segui, J. (1981). The syllable's role in speech segmentation. *Journal of Verbal Learning and Verbal Behavior, 20*(3), 298–305. https://doi.org/10.1016/S0022-5371(81)90450-3.

Mersad, K., & Dehaene-Lambertz, G. (2015). Electrophysiological evidence of phonetic normalization across coarticulation in infants. *Developmental Science, 19*(5), 710–722.

Mesgarani, N., Cheung, C., Johnson, K., & Chang, E. F. (2014). Phonetic feature encoding in human superior temporal Gyrus. *Science, 343*(6174), 1006–1010. https://doi.org/10.1126/science.1245994.

Moon, C., Lagercrantz, H., & Kuhl, P. K. (2013). Language experienced in utero affects vowel perception after birth : A two-country study. *Acta Paediatrica, 102*(2), 156–160. https://doi.org/10.1111/apa.12098.

Morais, J., Cary, L., Alegria, J., & Bertelson, P. (1979). Does awareness of speech as a sequence of phones arise spontaneously? *Cognition, 7*(4), 323–331. https://doi.org/10.1016/0010-0277(79)90020-9.

Pierrehumbert, J. B. (2003). Phonetic diversity, statistical learning, and Acquisition of Phonology. *Language and Speech, 46*(2–3), 115–154. https://doi.org/10.1177/00238309030460020501.

Räsänen, O., Doyle, G., & Frank, M. C. (2018). Pre-linguistic segmentation of speech into syllable-like units. *Cognition, 171*, 130–150. https://doi.org/10.1016/j.cognition.2017.11.003.

Richards, J. E. (1997). Effects of attention on infants' preference for briefly exposed visual stimuli in the paired-comparison recognition-memory paradigm. *Developmental Psychology, 33*(1), 22–31. https://doi.org/10.1037/0012-1649.33.1.22.

Swoboda, P. J., Kass, J., Morse, P. A., & Leavitt, L. A. (1978). Memory factors in vowel discrimination of Normal and at-risk infants. *Child Development, 49*(2), 332–339. https://doi.org/10.2307/1128695.

Trehub, S. E. (1973). Infants' sensitivity to vowel and tonal contrasts. *Developmental Psychology, 9*(1), 91–96. https://doi.org/10.1037/h0034999.

Trehub, S. E. (1976). The discrimination of foreign speech contrasts by infants and adults. *Child Development, 47*(2), 466–472. https://doi.org/10.2307/1128803.

Tsurumi, S., Kanazawa, S., Yamaguchi, M. K., & Kawahara, J. (2019). Attentional blink in preverbal infants. *Journal of Vision, 19*(10), 108b. https://doi.org/10.1167/19.10.108b.

Valenza, E., & Mento, G. (2016). Spatiotemporal neurodynamics of automatic temporal expectancy in 9-month old infants. *Scientific Reports, 6*, 36525. https://doi.org/10.1038/srep36525.

Werker, J. F., & Tees, R. C. (1984). Cross-language speech perception : Evidence for perceptual reorganization during the first year of life. *Infant Behavior and Development, 7*(1), 49–63.

Werker, J. F., & Tees, R. C. (1999). Influences on infant speech processing : Toward a new synthesis. *Annual Review of Psychology, 50*(1), 509–535.

Werker, J. F., & Tees, R. C. (2005). Speech perception as a window for understanding plasticity and commitment in language systems of the brain. *Developmental Psychobiology, 46*(3), 233–251. https://doi.org/10.1002/dev.20060.

Yeung, H. H., & Werker, J. F. (2009). Learning words' sounds before learning how words sound : 9-month-olds use distinct objects as cues to categorize speech information. *Cognition, 113*(2), 234–243.