

# SCIENTIFIC REPORTS



OPEN

## Enhanced Neonatal Brain Responses To Sung Streams Predict Vocabulary Outcomes By Age 18 Months

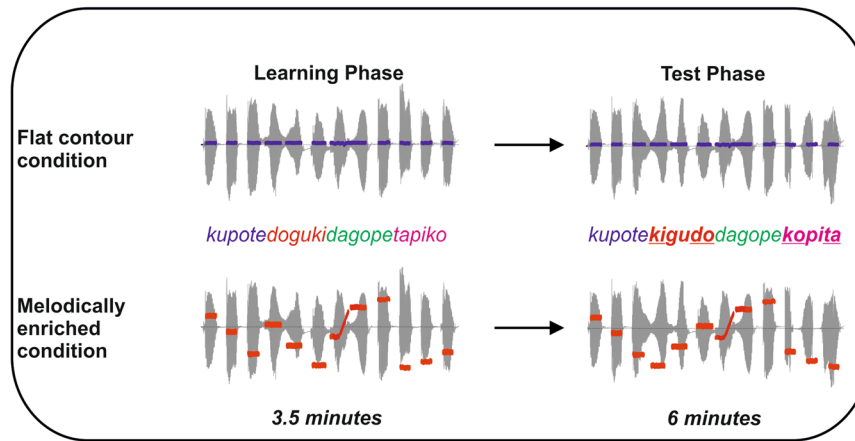
Clément François<sup>1,2,3</sup>, Maria Teixidó<sup>1</sup>, Sylvain Takerkart<sup>4</sup>, Thaïs Agut<sup>3,5</sup>, Laura Bosch<sup>1,2,3,6</sup> & Antoni Rodriguez-Fornells<sup>1,2,7</sup>

Words and melodies are some of the basic elements infants are able to extract early in life from the auditory input. Whether melodic cues contained in songs can facilitate word-form extraction immediately after birth remained unexplored. Here, we provided converging neural and computational evidence of the early benefit of melodies for language acquisition. Twenty-eight neonates were tested on their ability to extract word-forms from continuous flows of sung and spoken syllabic sequences. We found different brain dynamics for sung and spoken streams and observed successful detection of word-form violations in the sung condition only. Furthermore, neonatal brain responses for sung streams predicted expressive vocabulary at 18 months as demonstrated by multiple regression and cross-validation analyses. These findings suggest that early neural individual differences in prosodic speech processing might be a good indicator of later language outcomes and could be considered as a relevant factor in the development of infants' language skills.

Very early in life human neonates already exhibit remarkable auditory learning capacities<sup>1,2</sup>. The ability to detect pitch changes is already functional prenatally as early as after 27 weeks of gestation<sup>3</sup> and prenatal exposure to musical pieces induces memory traces that are still present 4 months after birth<sup>4</sup>. The early ability to process auditory stimuli during the last three months of gestation is associated to the fast migration of neurons from the ventricular zone to the cortical plate which may already determine the specific columnar organization of sensory areas observed in the left and right auditory cortices<sup>5,6</sup>. Interestingly, while human newborns already activate a left-lateralized fronto-temporal network including cortical perisylvian regions (superior temporal, inferior frontal and inferior parietal cortices) in response to linguistic stimuli<sup>7-9</sup>, a right lateralized activation is observed for voice or music processing<sup>10,11</sup>. The different patterns of activations observed for specific features of the auditory input may well be explained by differences in temporal sensitivity of the left and right auditory cortices<sup>12,13</sup>. Interestingly, two-day-old newborns can also build up temporal expectations and pick up the rhythmical regularities in the auditory input<sup>14</sup>. Furthermore, newborns show sensitivity to the rhythmic properties of speech and music<sup>15,16</sup> and their cry melodies have been found to be already language specific<sup>17</sup>. In this context, parents' use of infant-directed speech (IDS) containing exaggerated prosodic properties has been found to be important in facilitating speech perception and language acquisition<sup>18-20</sup>.

One of the earliest problems infants face in language learning is breaking the continuous speech input into word-form units, which is a necessary step in language acquisition. As opposed to written language, speech unfolds in time and no systematic cues signal word boundaries. One of the possible mechanisms proposed to solve this segmentation problem relies on humans' ability to track the statistical regularities present in the speech signal, based on the computation of transitional probabilities between adjacent syllables<sup>21</sup>. Using behavioural

<sup>1</sup>Cognition and Brain Plasticity Group, Bellvitge Biomedical Research Institute IDIBELL, L'Hospitalet de Llobregat, Barcelona, Spain. <sup>2</sup>Department of Cognition, Development and Educational Psychology, University of Barcelona, Barcelona, Spain. <sup>3</sup>Institut de Recerca Pediàtrica Hospital Sant Joan de Déu, Barcelona, Spain. <sup>4</sup>Aix Marseille Univ, CNRS, INT, Inst Neurosci Timone, Marseille, France. <sup>5</sup>Department of Neonatology, Hospital Sant Joan de Déu, Barcelona, Spain. <sup>6</sup>Institut de Neurociències, University of Barcelona, Barcelona, Spain. <sup>7</sup>Institució Catalana de Recerca i Estudis Avançats, ICREA, Barcelona, Spain. Correspondence and requests for materials should be addressed to C.F. (email: [fclement24@hotmail.com](mailto:fclement24@hotmail.com))



**Figure 1.** After a learning phase of 3.5 minutes (left side, learning phase), an implicit test (right side, test phase) was performed in which illegal word-forms violating the statistical structure (e.g. a legal ABC word-form in the language would become a CBA illegal word-form) appeared at pseudo-random positions in the stream. In the flat contour condition, all the syllables were spoken on the same pitch, thus resulting in a monotonous stream of syllables in which the only cue to word segmentation were the transitional probabilities between adjacent syllables. In the melodically enriched condition, each syllable was associated to a unique pitch, thus resulting in a continuous stream of sung syllables in which transitional probabilities and pitch modulations could be used for word segmentation. The blue traces represent the constant pitch used in the flat contour condition for both learning and test phases. The red traces represent the varying pitches used in the melodically enriched condition for both learning and test phases. In both conditions, the syllable duration was set to 350 ms thus leading to 1050 ms tri-syllabic pseudo-words.

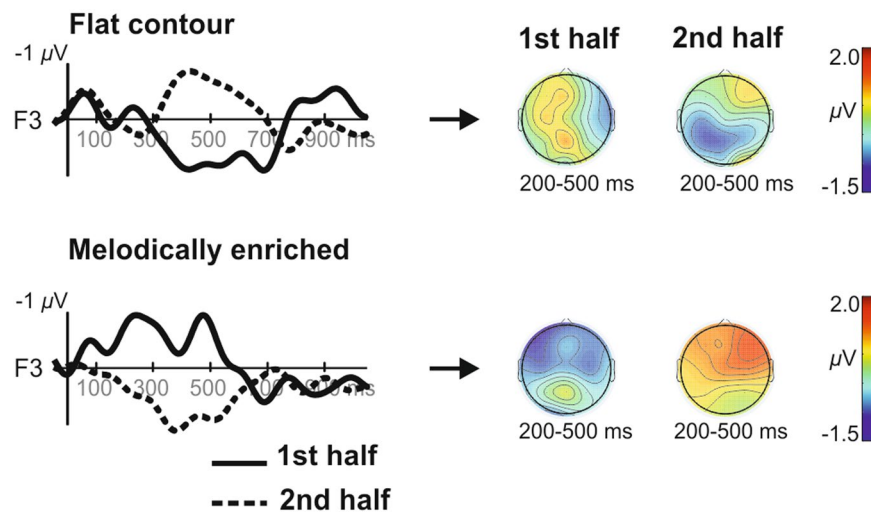
measures, seven- to nine-month-old infants have been shown to use statistical information to extract word-form units from speech streams<sup>22</sup>. Recent work has also revealed newborns' successful statistical learning in the visual, auditory and speech domains<sup>23–25</sup>.

Prosodic cues such as lexical stress or pauses have also been shown to be crucial for successful word segmentation in both adults<sup>26,27</sup> and infants<sup>22,28</sup>. Interestingly, a recent EEG study has provided the first experimental evidence showing that IDS may facilitate statistical learning in sleeping neonates. This facilitation was reflected by enhanced brain responses for IDS as compared to adult-directed speech (ADS) material and thus confirmed the benefit of prosodic cues on early word-form segmentation<sup>29</sup>.

Following this idea, we investigated the benefit of prosodic cues on word-form segmentation in newborns by directly comparing brain activity elicited by flat or melodically enriched syllable sequences of connected speech of an artificial language in a within-subject design. With this aim, we developed an innovative procedure in which following an initial learning phase, newborns were exposed to an implicit test phase in which statistically illegal word-forms were pseudo-randomly inserted. Crucially, the brain activity collected in our paradigm allowed us to study both learning-related electrophysiological modulations (learning phase) and the outcome of the learning process (implicit test phase). The presence of specific neural signatures associated to the detection of structural violations (i.e. illegal word-forms) may indirectly reflect the strength of the underlying word-form memory traces created during the preceding learning phase<sup>30,31</sup>.

We measured event-related brain potentials (ERPs) in 28 healthy 2- to 4-day-old sleeping human neonates exposed to continuous speech streams composed of four tri-syllabic items with and without pitch contour modulations (“flat contour condition” and “melodically enriched condition”; see Fig. 1, for experimental design). Considering recent data obtained with IDS material<sup>29</sup>, our working hypothesis was that melodic cues consistently coupled with statistical information in the sung streams would enhance speech segmentation in neonates. This benefit should be reflected during the learning phase by different dynamics of the ERP components elicited by melodically enriched and flat contour speech streams. Considering previous studies reporting the modulation of either a fronto-central positivity or a fronto-central negative ERP component in the 200–500 ms latency band post word-onset during speech segmentation tasks both in adults<sup>30–33</sup> and neonates<sup>25,29</sup>, we focused our analyses in the ERP components that developed during this time-window. For the analyses of the implicit test phases, we capitalized on previous research showing that neonates can detect low level<sup>34</sup> as well as more complex/abstract changes in the auditory inputs<sup>35</sup>. These studies have consistently shown that deviant stimuli rarely occurring in a sequence of standard stimuli elicited a mismatch response (MMR), considered as an electrophysiological index of pre-attentive, implicit auditory processing and may reflect the formation of an echoic memory trace within the auditory cortex<sup>36</sup>. Here, we included illegal items containing two syllabic order changes, one in the first and the other in the last syllable position. Considering previous studies using similar types of deviants<sup>30,31</sup>, we expected to observe enhanced neural signatures of illegal word-form detection, indexed by the presence of a MMR elicited by the first and second syllabic order changes, developing around 150 ms after syllable onset, in the melodically enriched condition when compared to the flat contour condition, indexing more efficient word-form extraction.

## LEARNING Phase



**Figure 2.** Time-course of the Learning effects. Grand averages ERPs at F3 electrode across 27 newborns recorded during each block (thick = 1<sup>st</sup> block, dotted = 2<sup>nd</sup> block) of the learning phases in both conditions (flat and melodically enriched). The isovoltage topographical maps show the distribution of the mean amplitude in the 200–500 ms latency band for each learning block.

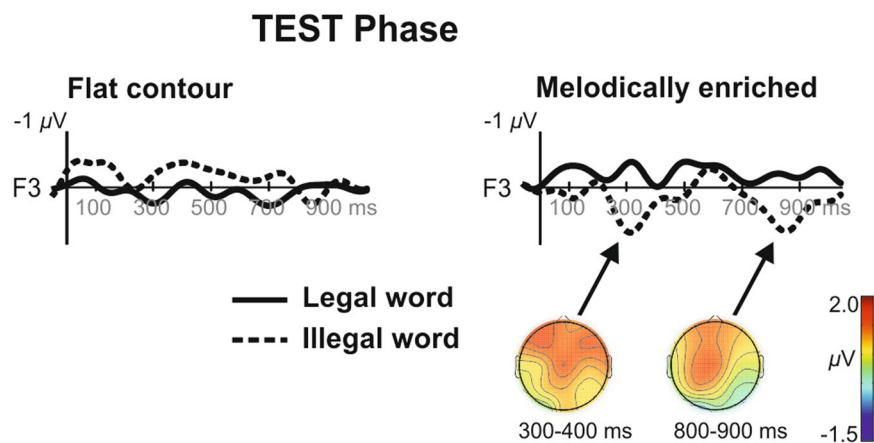
Interestingly, previous studies have demonstrated a link between behavioural and electrophysiological indices of early speech discrimination abilities for native and non-native phonemes and later language outcomes (expressive and receptive vocabulary)<sup>37</sup>. In addition, previous research on early word segmentation skills using natural speech stimuli have also shown that behavioural and electrophysiological data obtained during speech segmentation tasks in infants ranging from 7.5 to 12 months of age predicted language outcomes later in development<sup>38–40</sup>. Based on these previous studies, we aimed to provide first evidence for the predictive power of neonatal EEG signatures of speech segmentation for later language outcomes, assessed from vocabulary measures at 18 months of age. With this aim, we longitudinally followed the neonates enrolled in the EEG study and gathered measures of expressive vocabulary with the MacArthur Bates Developmental Inventory (MCDI) as well as the language subscale scores from the Bayley Scales of Infant Development (BSID-III) at 18 months of age. In order to explore the link between EEG signatures of early speech segmentation and expressive vocabulary at 18 months, we used Pearson correlations, linear multiple regression and cross-validation procedures with the EEG features reflecting the learning brain dynamics (for both flat contour and melodically enriched conditions), as well as the significant MMR found during the test phases of the melodically enriched condition, as predictive indices. Specifically, we hypothesized that neonatal brain sensitivity to melodically enriched speech streams could be predictive of expressive vocabulary at 18 months of age.

## Results

**Learning phases.** The EEG data from the learning phases were divided in two non-overlapping blocks for both conditions. This was done in order to track the dynamic evolution of the ERPs through exposure to the streams (see Fig. 2 and Supplementary Fig. 1). The results of the repeated-measures analysis of variance (ANOVA) with condition (flat contour and melodically enriched), blocks (block 1 and block 2) and electrode (16 levels) as within-subjects factors showed different patterns of ERP modulation as a function of exposure as revealed by a significant condition by block interaction ( $F(1,25) = 5.02$ ;  $P = 0.03$ ;  $\eta_p^2 = 0.17$ ). In the flat contour condition, a clear broadly distributed positivity was observed in the first block while a central-parietal negativity was found during the second block. Nonetheless, the between blocks pairwise comparison performed on the mean amplitude across all electrodes showed no significant difference between blocks (1<sup>st</sup> block:  $0.35 \mu\text{V} \pm 1.41$ ; 2<sup>nd</sup> block:  $-0.11 \mu\text{V} \pm 1.72$ , post-hoc test:  $P = 0.32$ ). By contrast, the melodically enriched condition elicited a fronto-central negative ERP component during the first block ( $-0.48 \mu\text{V} \pm 1.65$ ) whereas a clear positivity was observed in the second block ( $0.88 \mu\text{V} \pm 2.36$ ). The between blocks post-hoc LSD tests performed on the mean amplitude across all electrodes in the melodically enriched condition showed significant differences (Post-hoc LSD test:  $P = 0.03$ ). The interaction condition by block by electrode was not significant, nor was the main effect of block or condition (all  $P$ 's  $> 0.3$ ). Despite visual differences in the scalp distribution of the negative components found in the different learning blocks (see isovoltage maps at Fig. 2, right side), a further comparison (repeated-measures ANOVA with the factors condition and electrodes) between the topography of both components showed no significant differences (main effect of condition  $F(1,25) = 0.60$ ;  $P = 0.44$ ;  $\eta_p^2 = 0.02$ ; condition by electrode interaction  $F(1,25) = 2.06$ ;  $P = 0.11$ ;  $\eta_p^2 = 0.74$ ).

Flat Contour	300–400						800–900					
	M	SD	<i>t</i>	<i>P</i>	<i>P<sub>FDR</sub></i>	Cohen's <i>d</i>	M	SD	<i>t</i>	<i>P</i>	<i>P<sub>FDR</sub></i>	Cohen's <i>d</i>
Fp1	−0.07	2.1	−0.2	0.87	0.94	0.02	−0.13	1.4	−0.5	0.64	0.88	0.19
Fp2	−0.19	2.2	−0.4	0.66	0.94	0.27	−0.24	1.1	−1.0	0.31	0.88	0.41
F4	0.06	1.9	0.1	0.88	0.94	0.19	−0.08	1.3	−0.3	0.76	0.88	0.12
Fz	0.03	1.8	0.1	0.94	0.94	0.08	−0.07	1.5	−0.2	0.80	0.88	0.09
F3	−0.38	2.3	−0.8	0.42	0.94	0.37	−0.17	1.2	−0.7	0.50	0.88	0.32
T7	−0.98	2.1	−2.3	0.02	0.20	0.8	−0.47	1.3	−1.9	0.07	0.70	0.76
C3	−0.34	2.6	−0.7	0.52	0.94	0.21	−0.21	1.7	−0.6	0.52	0.88	0.26
Cz	−0.21	3.1	−0.3	0.74	0.94	0.31	−0.05	1.8	−0.1	0.89	0.89	0.06
C4	−0.04	2.3	−0.1	0.94	0.94	0.06	0.12	1.5	0.4	0.67	0.88	0.17
T8	0.66	2.2	1.6	0.13	0.65	0.05	0.09	1.1	0.4	0.70	0.88	0.15
Melodically Enriched	300–400						800–900					
	M	SD	<i>t</i>	<i>P</i>	<i>P<sub>FDR</sub></i>	Cohen's <i>d</i>	M	SD	<i>t</i>	<i>P</i>	<i>P<sub>FDR</sub></i>	Cohen's <i>d</i>
Fp1	1.27	2.3	2.9	0.008	0.07	1.1	1.49	2.0	3.8	<b>0.0008</b>	0.003	1.5
Fp2	1.33	2.8	2.4	0.02	0.07	0.9	1.53	2.0	3.9	<b>0.0006</b>	0.003	1.5
F4	1.13	2.9	1.9	0.06	0.10	0.8	1.17	1.9	3.1	<b>0.004</b>	0.01	1.2
Fz	1.06	3.2	1.7	0.10	0.14	0.7	1.15	2.1	2.8	<b>0.009</b>	0.01	1.1
F3	1.29	3.0	2.2	0.03	0.07	0.9	1.32	2.3	2.9	<b>0.007</b>	0.01	1.2
T7	0.77	0.5	1.6	0.12	0.15	0.6	0.48	1.9	1.3	0.2	0.20	0.5
C3	1.10	2.7	2.1	0.04	0.08	0.8	1.02	2.0	2.6	<b>0.01</b>	0.01	1.0
Cz	1.40	3.2	2.2	0.03	0.07	0.9	1.38	1.9	3.7	<b>0.001</b>	0.003	1.5
C4	0.69	2.9	1.2	0.24	0.26	0.5	0.86	2.0	2.2	0.04	0.05	0.9
T8	0.52	0.6	0.9	0.39	0.39	0.3	0.57	1.7	1.7	0.10	0.11	0.6

**Table 1.** Test phase: Results of the *t*-tests comparing the difference waveform (illegal minus legal words) against zero for the two time-windows (300–400 and 800–900 ms) and for the 10 electrodes in the two conditions. The significant differences remaining significant after FDR correction ( $n = 10$  comparisons;  $P < 0.05$ ) are highlighted in bold.



**Figure 3.** Grand averages ERP at F3 electrode across 26 newborns for legal and illegal word (thick = legal, dotted = illegal words) of the test phases in both conditions. The topographical maps show the distribution of the mean amplitude for illegal words in the significant time-windows.

**Test phases.** During the test phases, and as a first step that aimed to detect the presence of a MMR in each condition, we compared the mean amplitude of the difference waveform (illegal minus legal words) against zero separately for each condition (see Table 1). Because during the test phases, the illegal words which were built by reversing the order of the first and last syllables composing each of the words (see Fig. 1), we expected to observe two MMR developing after the first and last syllables (see ERP results in Fig. 3, and Supplementary Fig. 2). Results of this analysis revealed no significant differences in the flat contour condition in neither of the electrodes and time-windows. By contrast, we found significant differences against zero in the melodically enriched condition in two time-windows (300–400 and 800–900 ms) with even more significant effects for the second time-window (see Table 1).

Following this first analysis, we then contrasted the ERPs elicited by statistically legal and illegal words separately in each condition. We compared the mean amplitudes for illegal and legal words in each time-windows (300–400 and 800–900 ms), corresponding to the activity elicited by each of these two syllabic order changes. In order to evaluate these effects, we performed a repeated-measure ANOVA including the factors item type (legal vs. illegal words) and electrode (16 levels) for each time window and condition separately. No significant differences were found in the flat contour condition for illegal words in neither of the two time-windows (1<sup>st</sup> time-window: item type:  $F(1,25) = 0.35$ ;  $P = 0.56$ ;  $\eta_p^2 = 0.01$ ; item type by electrode:  $F(1,25) = 1.08$ ;  $P = 0.36$ ;  $\eta_p^2 = 0.04$ ; 2<sup>nd</sup> time-window: item type:  $F(1,25) = 0.16$ ;  $P = 0.69$ ;  $\eta_p^2 = 0.01$ ; item type by electrode:  $F(1,25) = 0.93$ ;  $P = 0.52$ ;  $\eta_p^2 = 0.04$ ). In contrast, for the melodically enriched condition, a clear positivity was found for illegal words occurring at the first and third syllable order violations (see Fig. 3). Both positive ERP components were significantly larger for illegal than for legal words (1<sup>st</sup> time-window: item type:  $F(1,25) = 4.81$ ;  $P = 0.03$ ;  $\eta_p^2 = 0.16$ ; item type by electrode:  $F(1,25) = 1.20$ ;  $P = 0.31$ ;  $\eta_p^2 = 0.05$ ; 2<sup>nd</sup> time-window: item type:  $F(1,25) = 13.16$ ;  $P = 0.001$ ;  $\eta_p^2 = 0.34$ ; item type by electrode:  $F(15,375) = 2.41$ ;  $P = 0.07$ ;  $\eta_p^2 = 0.08$ ). Because of the almost significant item type by electrode interaction for the second time-window, we performed post-hoc LSD tests and found that illegal words elicited larger positive ERP component than legal words over the 8 fronto-central electrodes (Fp1, Fp2, F3, Fz, F4, C3, Cz, C4; all  $P$ 's < 0.01).

We also directly compared the difference waveforms (illegal minus legal words) between the two conditions separately for the first and second time-windows. These difference waveforms were submitted to a repeated-measures ANOVA with the factor condition (flat contour vs. melodically enriched) and electrode (16 levels). Compared to the flat contour condition, we found significantly enhanced brain responses in the melodically enriched for the first (condition:  $F(1,25) = 4.33$ ;  $P = 0.04$ ;  $\eta_p^2 = 0.15$ ; condition by electrode,  $F(15,375) = 1.21$ ;  $P = 0.31$ ;  $\eta_p^2 = 0.05$ ) and the second time-window (condition:  $F(1,25) = 9.34$ ;  $P = 0.005$ ;  $\eta_p^2 = 0.27$ ; condition by electrode:  $F(15,375) = 2.36$ ;  $P = 0.06$ ;  $\eta_p^2 = 0.09$ ).

**Cognitive development and language outcomes at 18 months.** Cognitive and language outcome measures could only be obtained from thirteen infants from the original sample (see Method section). The mean productive vocabulary size from the MCDI was 48.8 words (Mdn = 38; percentile 35; SD = 45.07). Cognitive and language attainment assessed with the BSID-III yielded mean standardized scores of 113 (Mdn = 115; SD = 10.07) and 102 (Mdn = 97; SD = 11.82) for the cognitive and the language subscales respectively. Importantly, both measures were within the normal range (100 +/- 15).

**Relationship between neonatal brain responses and expressive vocabulary at 18 months.** We first explored the correlations between the relevant EEG measures gathered at 2 days-old in the segmentation task (learning and test phases) and the measures of expressive vocabulary at 18 months (raw MCDI scores and the language scores of the BSID-III). Specifically, considering previous ERP results described above, we selected four EEG measures as possible predictors: (i) the learning brain dynamics reflected by the difference waveforms (block 2 minus block 1) of the learning phases for both the flat and melodically enriched conditions (L\_flat and L\_melody) in the 200–500 ms time-window, and (ii) for the melodically enriched condition, the difference waveforms obtained by subtracting illegal minus legal word-forms conditions for the two syllabic order changes (300–400 and 800–900 ms time-windows: TW1\_melody and TW2\_melody).

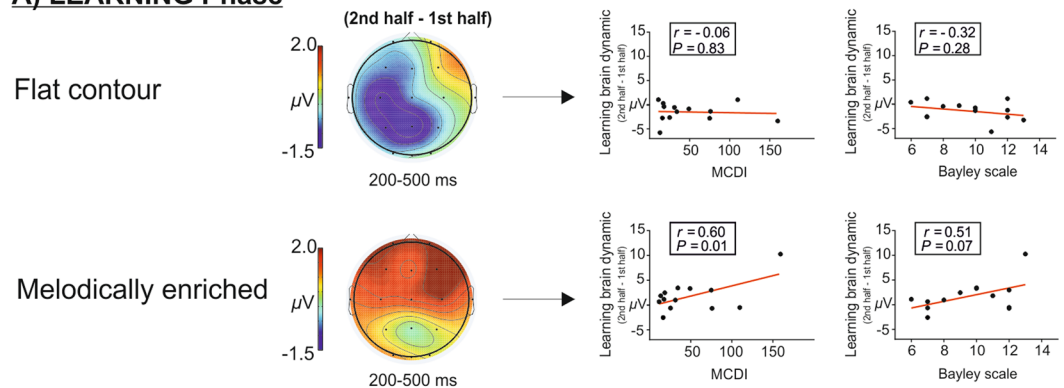
For the ERP variables of the learning phases, the correlation between the MCDI score with the learning effect in the melodically enriched condition (L\_melody) was almost significant after FDR correction ( $r(11) = 0.60$ ,  $P_{FDR} = 0.06$ ). This was not the case in the flat contour condition (see Fig. 4). These results were confirmed by a marginally significant correlation with the language subscale score of the BSID-III ( $r(11) = 0.51$ ,  $P_{FDR} = 0.14$ ; see Fig. 4) in the melodically enriched condition and not in the flat contour one.

For the two ERP variables of the test phase, we found a significant correlation after FDR correction between the second syllabic order change (TW2\_melody) and the MCDI scores (MCDI:  $r(11) = -0.60$ ,  $P_{FDR} = 0.04$ , see Fig. 4). This result was confirmed by a marginally significant correlation with the language subscale score of the BSID-III ( $r(11) = -0.59$ ,  $P_{FDR} = 0.06$ ).

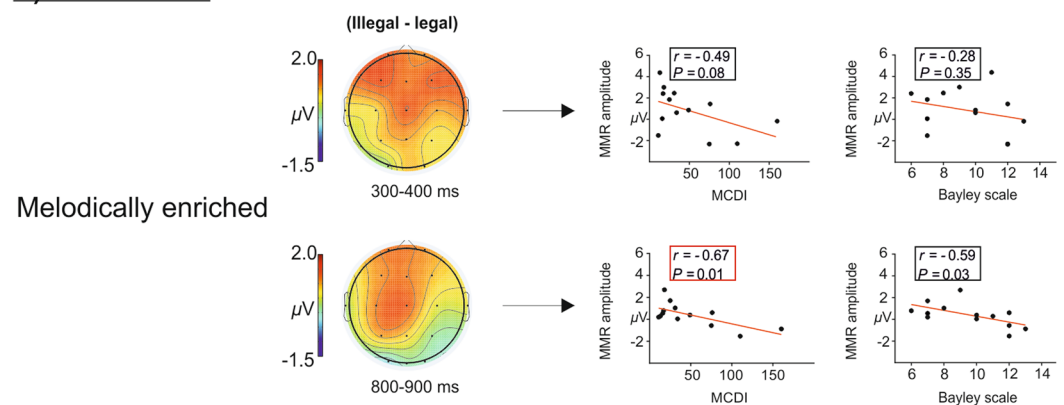
As a second step, we performed a backward multiple regression analysis introducing the four ERP measures as possible predictors of the MCDI scores<sup>41</sup>. In the first model with the four ERP measures included, the predictors accounted for 64% of the variance in expressive vocabulary ( $R^2 = 0.64$ ;  $F(4,8) = 3.59$ ,  $P = 0.05$ ). The final model retained only two predictors from the melodically enriched condition: the learning effect and the second violation effect. This model explained 63% of variance and was highly significant ( $R^2 = 0.63$ ;  $F(2,10) = 8.38$ ,  $P = 0.007$ ; see Table 2 and Fig. 4).

**Cross-validation and generalization.** Finally, we examined the generalizability of the results obtained in the multiple regression analysis using a cross-validation strategy based on a machine learning evaluation framework (see Methods). Specifically, we retained the two significant EEG features – the learning brain dynamic (block 2 minus block 1) and the second violation effect (illegal minus legal words) of the melodically enriched condition – in order to further quantify the out-of-sample generalization power of such multiple regression as a predictive model of the MCDI score. Our leave-two-subjects-out cross-validation yielded 156 predicted MCDI scores, which we compared to the true scores. The two complementary evaluation criteria that we computed ( $R^2$  of the linear regression, 0.56, and mean absolute error, 34.06) were found to be significant when compared to their empirical null distribution estimated using 5000 permutations ( $P = 0.046$  and  $P = 0.042$  respectively). This analysis directly demonstrates the potential of these two EEG features as biomarkers of future acquired vocabulary.

### A) LEARNING Phase



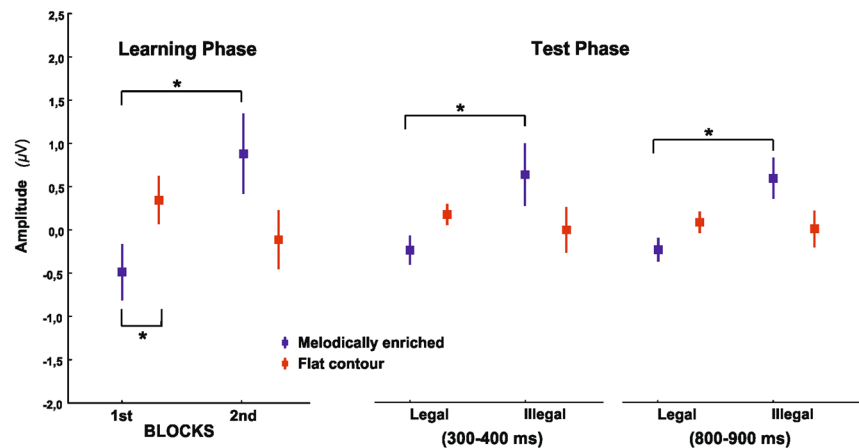
### B) TEST Phase



**Figure 4.** Link between neonatal brain responses in the learning and test phases and expressive vocabulary measures obtained from the MCDI and the Bayley-III language subscale. **(A)** the scatter plots show the correlations between both the raw MCDI scores and the language subscale score of the Bayley-III and the learning brain dynamics averaged across all channels and for each condition. A marginally significant correlation is observed between MCDI scores and the learning brain dynamics in the melodically enriched condition after FDR correction ( $n = 4$  comparisons,  $P < 0.05$ ). **(B)** the scatter plots show the correlations between the raw MCDI scores and the Bayley-III language subscale score and the mean amplitude of the difference waveform (illegal minus legal words) for the two time-windows averaged across all EEG channels. The correlation between the MCDI scores and the second MMR in the melodically enriched condition survives the FDR correction ( $n = 4$  comparisons,  $P < 0.05$ ) and is highlighted in red.

Model	Unstandardized coefficients		Standardized coefficients	t	P
	B	SE	$\beta_k$		
<i>Model 1</i>					
(Constant)	49.098	13.378		3.595	0.007
L_flat	0.212	5.844	0.009	0.036	0.97
L_melody	6.684	3.291	0.465	2.031	0.07
TW1_melody	-3.822	7.503	-0.174	-0.509	0.62
TW2_melody	-17.552	13.821	-0.421	-1.270	0.24
<i>Final Model</i>					
(Constant)	47.575	10.800		4.405	0.001
L_melody	6.263	2.896	0.436	2.163	0.05
TW2_melody	-22.929	8.403	-0.550	-2.729	0.02

**Table 2.** Coefficients for the multiple regression with the four neonatal EEG features as predictors ( $N = 13$ ). The abbreviations L\_flat and L\_melody correspond to the learning brain dynamics (block 2 minus block 1) in the flat contour and melodically enriched condition respectively. Viol\_melody1 and 2 correspond to the difference waveform mean amplitude in the first and second time-windows for the melodically enriched condition (see Methods).



**Figure 5.** Results of the learning phases (Left): Learning modulations of the ERP mean amplitude in the 200–500 ms latency band for the flat contour (red bars) and melodically enriched conditions (blue bars). Bars denote average amplitudes across all channels (with SEMs) in the two learning blocks. An opposite pattern of ERP modulation is observed in the melodically enriched as compared to the flat contour condition. Results of the test phases (Right): Bars denote average brain response in the 300–400 and 800–900 ms latency band (with SEMs) for legal and illegal word. Significant differences were found in the melodically enriched condition only (blue bars).

## Discussion

The present study reveals in newborns the electrophysiological brain dynamics during statistical learning of flat contour and melodically enriched speech streams. Our results indicate that during the learning phase, melodically enriched and flat contour streams elicited different ERP modulations through exposure. While flat contour streams elicited a positivity during the first block followed by a negativity during the second block, the opposite pattern was observed for the melodically enriched condition. Importantly, brain responses obtained during the implicit test phase revealed successful detection of statistical violations in the melodically enriched but not in the flat contour condition (see Fig. 5). Moreover, using a longitudinal design, we provide empirical and computational evidence for a link between the enhanced brain responses to melodically enriched streams (prosodic modulations) and later expressive vocabulary measures at 18 months.

Previous studies on the building of early word-form memory traces in 6- to 20-month-old infants have consistently revealed a frontal negativity in the 200–500 ms time-window whose amplitude increased with word familiarity. This frontal component has been taken to reflect the emergence of word-form memory traces in infants<sup>40,42</sup>. Additionally, adult studies on speech segmentation have delineated the brain dynamics occurring during the learning phases<sup>30–33</sup>. Results of these studies have consistently revealed modulations of both a P2 and a fronto-central N400-like component through exposure to the streams. Taken together, our results extend previous research<sup>25</sup> by showing that newborns can already extract word-forms from continuous speech based on statistical cues, as they already exhibit the characteristic brain response indexing the emergence of proto-lexical memory traces<sup>43</sup>. Interestingly, the learning brain dynamics observed in adult participants are tightly linked to the behavioural recognition of the word-forms embedded in the speech streams: results have shown that good learners exhibit a fast increase of the fronto-central N400-like amplitude after two minutes, while poorer learners showed a slower increase of this negativity through exposure<sup>30,32</sup>. In neonates, different patterns of positive and negative ERP modulation have been observed during exposure to the statistical auditory streams<sup>24,25,29</sup>. Here, we observed a fronto-central negative component during the first half of the learning phase followed by a broad positivity during the second half in the melodically enriched condition. Surprisingly, an opposite pattern of ERP modulation was found in the flat contour condition with a positivity followed by a negativity through stream exposure. Nonetheless, the longitudinal design used here provides additional evidence for the importance of the learning brain dynamics during speech segmentation. Indeed, the only significant relationships between neonatal brain responses and expressive vocabulary at 18 months were found for the melodically enriched condition only (see Table 2). The predictive power of these potential biomarkers related to language abilities was further confirmed by the machine learning approach we used here. Therefore, considering previous results on speech segmentation in adults, as well as results on word familiarity in infants showing similar negativities at similar latencies<sup>25,40,42</sup>, our results suggest that melodic information does indeed facilitate speech segmentation in neonates by enabling a more efficient extraction of word-forms.

However, none of the previous studies we have discussed have assessed the ability to detect on-line statistical violations after learning had taken place. This is crucial because during the learning phase of such statistical learning paradigms, only sufficiently reinforced items may survive interference and be stored in long-term memory<sup>44</sup>. Considering this constraint and the fact that behavioural measures are difficult to collect at a very early age (in 2-day-old infants), we sought neural evidence of online detection of structural violations that may indicate the building of sufficiently robust word-form memory traces. During the implicit test phase, illegal word-forms induced enhanced brain responses for the violations occurring at the first and last syllable positions in the

melodically enriched condition only. No significant difference was encountered in the flat contour condition. While a fronto-central negative MMR has been observed in response to similar statistical violations in adults<sup>30</sup>, a positive MMR has been systematically reported in newborns for different types of auditory deviants including simple and complex pitch regularities<sup>45,46</sup>. This component is also observed for speech sound changes<sup>1,47</sup> and its amplitude predicts later language development<sup>48,49</sup>. Taken together, the results of the test phase confirmed that melodically enriched streams facilitated the building of robust word-form memory traces and triggered strong enough expectations to successfully detect structural violations of the extracted word-forms in the test phase. In addition, these results further confirmed the importance of brain dynamics observed during the learning phase, as neonates were sensitive to changes in the statistical structure only for the melodically enriched streams. In sum, while brain signatures of successful speech segmentation were observed during the learning phase in both conditions (albeit with different dynamics), successful retention (i.e. the building of memory traces) of the extracted word-forms only occurred in the presence of melodic cues (as revealed in the test phase). Therefore, the intriguing question of why musical cues can boost speech segmentation needs to be addressed.

Compared to adult-direct speech, infant-directed speech is notably characterized by a slower rate of speech, a higher fundamental frequency and by the repetition of intonation structures<sup>50,51</sup>. In the present study, the repetition of the 3-tone pitch patterns in the melodically enriched streams contained systematic pitch-contour and pitch interval changes at word boundaries that may have aided the grouping of syllables<sup>52</sup>. Additionally, the syllables and pitches were correlated in the melodically enriched condition which may explain the enhanced brain responses for illegal words in the test phase. This would be in line with the observation of faster learning when a second source of information can provide redundant cues to identify the structure of the input<sup>53</sup>. Future studies are needed to determine whether newborns used both cues or deviants in pitch pattern alone. An alternative explanation might be related to the fact that melodically enriched streams, which contain pitch contour variations, have been found to enhance overall attention and emotion in infants<sup>50,51,54–56</sup>. For instance, compared to adult-directed speech stimuli, young infants prefer to listen to IDS probably due to its higher capacity to maintain infants' attention or convey emotional information<sup>54,56</sup>. Considering the similarities between infant-directed speech and music in terms of pitch and tempo modulations, as well as the impressive capacity of music to convey emotions<sup>57,58</sup>, songs could act in neonates as an innate catalyst of exogenous attention and emotion. In this context, it is well known that memory encoding is facilitated with high levels of attention and emotional processing<sup>59,60</sup>. This hypothesis is also consistent with behavioural data showing unsuccessful learning in infants when exposed to an impoverished emotional infant-directed speech<sup>61</sup>. Therefore, the presence of melodic cues in our experiment could have induced stronger attentional or reward-related processes promoting word-form extraction through the activation of reinforcement learning mechanisms. Indeed, the important role of emotion and reward in word learning and auditory learning in general has been recently proposed<sup>62,63</sup>. Moreover, it has also been observed that adults and infants could associate novel word-forms to objects more easily when presented with material containing the rich prosodic characteristics of IDS rather than with ADS<sup>18–20</sup>. Interestingly, it has been recently demonstrated that 12- and 24-months old infants' brain responses of word recognition is mediated by increased attention allocation through predictive coding processes<sup>64</sup>. Besides, music processing *per se* as well as the rewarding experience of music may also rely on predictive coding mechanisms<sup>57</sup>. Therefore, increased attentional and reward-related processing during exposure to melodically enriched streams might explain, at least in part, the facilitating effects observed in our data.

Nonetheless, the present study shows several limitations that must be mentioned. Further studies are needed to determine whether increasing the duration of the exposure phase in the flat contour condition may induce a similar dynamic evolution of the amplitude of the fronto-central negativity, as it is the case in the melodically enriched condition. Besides, it is important to note that the melodically enriched streams used here contained syllables that were “sung” in a musical structure with 11 notes out of 12 belonging to the C major scale, which may have induced the perception of a tonal centre<sup>65</sup> and could also have consequences on the formant information (e.g. F1, F2, and F3) which could be additional information to word segmentation. Indeed, newborns are sensitive to western musical chords and exhibit a large positive MMR to dissonant chords in a sequence of chords from the C major scale<sup>46</sup>. Further studies using different musical scales thus needed to determine the role of the tonality in triggering word-form extraction in early infancy as well as the role of formant information. Another limitation of the present study concerns the electrophysiology method and the use of the average mastoids as reference. In the present study, we performed an off-line referencing of our EEG data to the average mastoids in order to make our results easily comparable with previous reports on speech segmentation in neonates<sup>25,29,64</sup> and adults<sup>26,27,30–33,65,66</sup>. As recent studies using simulated and experimental data have suggested that the use of infinite reference based on the reference electrode standardization technique can avoid scalp topography distortions and localization errors<sup>67,68</sup>, future studies in neonates should compare the different methods of re-referencing. Another limitation to acknowledge concerns the comparisons of the two conditions, flat contour and melodically enriched. As compared to the flat contour streams, the melodically enriched streams contained pitch modulations that may have contaminated the ERP results, specifically when it comes to compare the difference waves between conditions. Moreover, the babies included in the present study did not present any hearing-related dysfunctions as all of them had normal results in the screening procedure performed during the first weeks of life using Automated Auditory Brainstem Responses. Data obtained about familiar history of dyslexia (only 86% of the families provided information on this issue), showed that the father of one of the infants had been diagnosed as dyslexic which is a factor known to influence early brain responses to speech sounds<sup>48,49</sup>. It will be important for future studies to take this parameter into account. Besides this, although an important percentage of the infants (62%) enrolled in the present study were mostly raised in monolingual families, some of them had been prenatally exposed to both Spanish and Catalan from their bilingual mothers. There is evidence showing differences in speech perception abilities between monolingual and bilingual infants early in development<sup>69–71</sup>. More crucial for the present study, sensitivity to differential language properties has been found to start even before birth in newborns whose mothers



had regularly spoken two rhythmically different languages during pregnancy<sup>72</sup>. Our bilingual infant participants had been prenatally exposed to two rhythmically close languages and evidence for robust intraclass language discrimination at birth is still lacking. However, an emerging early sensitivity to specific rhythm properties of the languages or attention to any differential features that might be captured before birth cannot be totally discarded in bilingual newborns, in contrast to those exposed prenatally to a single language. These differences may certainly start the last trimester of gestation when the cortical plate is setting up<sup>5,6</sup>. Therefore, it will be important to take these language exposure parameters into account in future studies.

Some practical considerations have also to be acknowledged. Firstly, we used a fixed order of presentation with the flat contour condition always presented first. This was done in order to reduce any possible facilitation effect carried by the melodically enriched condition. Moreover, we used different languages built with different CV syllables with no repetition of the same syllable across conditions to control for possible effect of interference between the two conditions. Even if we cannot clearly rule-out any carry-over effect from the first to the second condition, there is evidence, at least in adults, that learning a first artificial language decreases the capacity to learn the second one<sup>73</sup>. Future studies are necessary to further disentangle the possible effect of presentation order. Finally, in spite of a longitudinal design, vocabulary measures at 18 months could only be gathered for 13 of the participants in the original sample, a factor that constrains the statistical power of our analyses. Nonetheless, results of the machine learning evaluation procedure used here clearly demonstrated the importance of neonatal brain sensitivity to melodically enriched as compared to flat contour streams in later vocabulary acquisition. Further longitudinal studies with larger samples minimizing the effects of the observed attrition rate in these studies are needed to provide further support for the current data and to clearly characterize the link between neonatal brain sensitivity to prosodic modulations and later language abilities.

In summary, we found brain signatures of successful word-form segmentation for both melodically enriched and flat contour streams in 2-day-old neonates. The brain dynamics were different in the two conditions but crucially, neural responses of successful detection of structural violations was found in the melodically enriched condition only, indicating that sung streams induced a more efficient learning. Importantly, the longitudinal design used here offers a unique opportunity based on empirical and computational evidence to establish a link between newborns' brain sensitivity to melodically enriched speech streams and expressive vocabulary at 18 months. Previous results in adults and children have shown that musical training enhances speech segmentation<sup>65,66</sup>. Moreover, musical intervention at 9 months of age enhances brain sensitivity to the temporal structure of music and speech<sup>74</sup>. Taken together, our results converge in showing the importance of musically enriched input in the early steps of language acquisition. These findings may thus have clear implications for the important role of music-based programs on language learning.

## Methods

**Participants.** Twenty-eight healthy full-term neonates (mean age at testing = 2.8 days, SD = 0.9, range: 1–4 days; 14 boys; mean birth weight = 3375 g; SD = 506.4, range: 2640–4420 g; the mean Apgar score obtained at 5 minutes of life was >8 and the mean Apgar score obtained at 10 minutes of life was >9; mean gestational age = 40 weeks; SD = 0.8, range: 37–41 weeks) were enrolled in this experiment. Note that the Apgar score is an objective tool that measures different signs of physiologic adaptation after birth. An Apgar score of 7 or more indicates that the baby is in good condition. Newborns were recruited and tested at Hospital Sant Joan de Déu, Barcelona, Spain. Only infants born from parents whose native language was either Spanish or Catalan participated in the study. Information about their native language, the language predominantly used at home and an estimate of the daily use of each of the ambient languages (i.e. Spanish and Catalan) by the mother in the last three months of pregnancy was obtained through an interview at the hospital room. Although all parents could speak and understand both Catalan and Spanish, the interview revealed that 62% of the families had a predominant language use (above 75% of regular use of only Spanish or Catalan), the remaining families showing a more clear bilingual status in the daily use of the two languages present in the community. Parents were informed and signed a consent form at the beginning of the experimental session. This study was carried out in accordance with the guidelines of the Declaration of Helsinki (BMJ 1991; 302: 1194) and approved by the Ethical Committee of Hospital Sant Joan de Déu (University of Barcelona; CEIC-PIC-69-13). All neonates passed a hearing screening test (ABBR, Automated Auditory Brainstem Response) and an examination by a neonatologist at the delivery ward.

**Stimuli.** *Learning streams.* Two languages (L1 and L2) were built using a set of 24 synthetic syllables that were combined to give rise to two sets of four tri-syllabic pseudo-words (L1: KUPOTE, DOGUKI, TAPIKO, DAGOPE; L2: TUGODE, PAGODI, DUGAKE, KATIPU). In both conditions, the syllable duration was set to 350 ms thus leading to 1050 ms tri-syllabic pseudo-words. In the flat contour condition, all the syllables had the same pitch of 216 Hz. In the melodically enriched condition and for each language, each of the 12 syllables was associated with a distinct tone. Therefore each word had a unique melodic contour (KUPOTE: D<sub>4</sub>C<sub>4</sub>G<sub>3</sub>, DOGUKI: D<sup>b</sup><sub>4</sub>A<sub>3</sub>E<sub>3</sub>, TAPIKO: B<sub>3</sub>E<sub>4</sub>F<sub>4</sub>, DAGOPE: C<sub>3</sub>D<sub>3</sub>F<sub>3</sub>). No pitch-contour changes occurred within the pseudo-words (2 pseudo-words contained a rising pitch-contour while 2 contained a falling pitch-contour). The mean pitch interval within pseudo-words was significantly different from the mean interval between the pseudo-words ( $P = 0.02$ ). Moreover, pitch-contour changes could be used to segment the stream, as this cue was consistent and no other acoustic cues were inserted at word boundaries. The four learning streams were built by a random concatenation of the four words (without repetition of the same item twice in a row). All the stimuli were synthesized using Mbrola (<http://tcts.fpms.ac.be/synthesis/mbrola.html>) with the Spanish es2 database and thus were only built with syllables that were part of the Spanish-Catalan phonetic repertoire. The transitional probability between the syllables forming a word was 1.0, while the transitional probability between syllables spanning word boundaries

was 0.33. Each word was repeated 50 times in the stream leading to 3.5 minutes continuous speech streams. Importantly, no repetitions of syllables were made across the two language streams (L1 and L2).

**Tests streams.** The test streams consisted of 66 repetitions of each pseudo-word from the corresponding language in which test items were pseudo-randomly inserted as done previous studies<sup>30,31</sup>. The test items were built by reversing the order of the syllables composing each of the words. The test items were TEPOKU, KIGUDO, KOPITA and PEGODA for L1 and DEGOTU, DIGOPA, KEGADU, PUTIKA for L2. The test items were inserted in the continuous stream but could not follow the word from which it was derived. The overall probability of the test items in the test streams was 25% (22 repetitions of each novel pseudo-words).

**Experimental procedure and data acquisition.** The entire recording session took place directly in the room of the maternity unit. During the experimental session, infants were lying asleep on their cribs while being presented with the auditory streams (learning phase and implicit test) of both conditions in a within-subject design. All infants were presented with the flat contour condition followed by the melodically enriched condition. Half of the infants were exposed to one language in the flat contour and to the other in the melodically enriched condition and this was counter-balanced across participants. At least one of the 2 parents was present during the session. Stimuli were presented at a 70-dB volume from one loudspeaker positioned in front of the infants at a distance of about 1 m from the infant's crib. EEG was recorded from 16 scalp electrodes (Biosemi ActiveTwo system, Amsterdam University) located at standard positions (International 10/20 system sites: Fp1, Fp2, F3, F4, T7, C3, C4, T8, P3, P4, O1, O2, Fz, Cz, Pz and Oz). The data of three newborns had to be discarded due to excessive movements inducing major EEG artifacts. The EEG was amplified by Biosemi amplifiers with a band-pass of 0–102.4 Hz and was digitized at 250 Hz.

**Vocabulary and cognitive measures.** At 18 months, parents of the participants were contacted again and asked to come to the Lab for a language and cognitive development assessment. Parents were requested to fill out the expressive vocabulary checklist from the Spanish version of the MacArthur-Bates Communicative Development Inventory<sup>75</sup> (MCDI). This questionnaire provided us with a parental estimate of expressive vocabulary size. Infants were also tested with the Bayley Scales of Infant Development<sup>76</sup> (BSID-III). Only the Cognitive and the Language subscales were used. A general measure of cognitive development was obtained to ensure that all participants in the sample fell within the normal range. The Language subscale score was obtained as a more standard measure of toddlers' attainment in both receptive and expressive language development. We expected the latter measure to positively correlate with the expressive vocabulary reports from the MCDI, thus limiting the presence of parental biases in the way they reported toddlers' spontaneous word productions. Both the BSID-III cognitive and language subscales and the MCDI questionnaire were administered in a room of the APAL Infant Lab at the Hospital Sant Joan de Déu. The data of thirteen 18 month-olds could be satisfactorily gathered.

**EEG data analyses.** All the EEG data were processed and analyzed using custom scripts programmed in Matlab (version R2008b) and with functions of the EEGLAB toolbox. For the learning phases, EEG data were re-referenced offline to the left and right mastoidal electrodes and were filtered offline using a 0.5–20 Hz bandpass filter (FIR, Hamming window, max stopband attenuation = −53 dB, Max passband deviation = 0.0063; as done in previous reports studying the EEG correlates of the learning phases<sup>24,25,30,32</sup>). Continuous EEG data were then split into epochs from −50 ms to 1050 ms from word onset, baseline-corrected and epochs containing external artifacts exceeding  $\pm 120 \mu\text{V}$  were automatically removed from analyses. ERP data were analyzed by computing the mean amplitudes between 200 and 500 ms post word-onset independently of the sleep stage of the newborns. This time window was selected based on visual inspection and previous findings with statistical learning paradigms in infants and adults<sup>24,25,32,33</sup>. We used repeated-measures analysis of variance (ANOVA) with condition (flat contour vs. melodically enriched), blocks (first vs. second, based on 2 consecutive non-overlapping time windows of 1'45", 100 trials each) and electrodes (16 levels) as within-subject factors.

For each ANOVA, Greenhouse-Geisser sphericity corrections were applied when appropriate. Partial-eta-squared ( $\eta^2$ ) was computed for the main effects and interactions with other factors. Post-hoc tests were conducted using the Fisher LSD method as done in<sup>25</sup>. Cohen's d effect sizes were also computed and reported when appropriate.

For the test phases, EEG data were re-referenced offline to the left and right mastoidal electrodes and were offline filtered from 1 to 20 Hz (as done in previous auditory MMR studies in newborns and children<sup>35,77–79</sup>). Continuous EEG data were then split into epochs from −50 ms to 1050 ms from word onset, baseline-corrected and epochs containing external artifacts exceeding  $\pm 120 \mu\text{V}$  were automatically removed from analyses. Firstly, in order to detect the presence of a MMR, two-tailed Student's *t*-tests were used to compare the mean amplitude of the difference waveform against zero in each condition. Based on the results of this first exploratory analysis and on previous studies using similar types of deviants<sup>30</sup>, ERP data were analyzed separately for each condition by comparing the mean amplitudes for illegal and legal words in two time-windows (300–400 and 800–900 ms) covering the activity elicited by the two statistical violations (see<sup>30</sup> for similar analyses in adults). Finally, in a third step analysis that aimed to directly compare the two conditions in the two time-windows of interest, the difference waveforms of the two conditions (illegal minus legal words) were submitted to a repeated-measures ANOVA with the factor condition (flat contour vs. melodically enriched) and electrode (16 levels). As done for the learning phases, Greenhouse-Geisser sphericity corrections were applied when appropriate. Partial-eta-squared ( $\eta^2$ ) was computed for the main effects and interactions with other factors. Post-hoc tests were conducted using the Fisher LSD method as done in<sup>25</sup>. Cohen's d effect sizes were also computed and reported when appropriate.

**Relationship between expressive vocabulary at 18 months and neonatal brain responses.** As a first exploratory step, we assessed the linear relationship between the EEG measures gathered during the learning phases and the measures of expressive vocabulary at 18 months (raw MCDI scores and the language scores of the BSID-III). Considering the ERP results of the learning phases, we used the learning brain dynamics reflected by the mean amplitude averaged across all EEG channels for the 200–500 ms time-window of difference waveforms (block 2 minus block 1) in the learning phases for both the flat and melodically enriched conditions (L\_flat and L\_melody). The Pearson correlations between the difference waveforms (block 2 minus block 1) of the learning phases relative to the raw MCDI scores and the language scores of the BSID-III were evaluated. The correlations were considered significant for *P*-values lower than 0.0125 (after FDR correction with *n* = 4 comparisons).

To investigate the relationship between the EEG measures gathered during the test phases and the measures of expressive vocabulary at 18 months we only considered the two MMR occurring at each syllabic violation (difference illegal minus legal words in the 300–400 and 800–900 ms time-windows averaged across all EEG channels) in the melodically enriched condition. This choice was driven by the absence of significant MMR in this condition only. The Pearson correlations between the difference waveforms obtained by subtracting illegal minus legal word-forms conditions for the two syllabic order changes (300–400 and 800–900 ms time-windows: TW1\_melody and TW2\_melody) relative to the raw MCDI scores and the language scores of the BSID-III were evaluated. The correlations were considered significant for *P*-values lower than 0.0125 (after FDR correction with *n* = 4 comparisons).

As a second step, we performed a backward multiple linear regression to explore the possible predictive value and further confirm the contribution of each of the different EEG features gathered at two days after birth on the language outcomes obtained at 18 months. Specifically, we focused on expressive vocabulary measures, a variable that has been reported to significantly correlate with the speech segmentation skill assessed prelexically<sup>38–40</sup>. We used four neonatal EEG features to predict the MCDI raw scores. For the learning phases of both conditions, we considered the mean amplitude averaged across all EEG channels for the 200–500 ms time-window of the difference waveform (block 2 minus block 1). For the test phase, because significant MMR were found in the melodically enriched condition only, we considered the two MMR occurring at each syllabic violation (difference illegal minus legal words in the 300–400 and 800–900 ms time-windows averaged across all EEG channels).

Following this analysis, we explicitly quantified the predictive power of the significant EEG features using a machine learning evaluation framework<sup>80</sup>. For this, we removed two subjects from the dataset, fitted the same model on the eleven other subjects and used it to obtain a predicted MCDI value from the EEG features of the two left-out subjects. This procedure was repeated throughout a cross-validation procedure<sup>81</sup> (leave-two-subjects-out) where all 78 combinations of two-out-of-thirteen newborns were used to obtain out-of-sample predicted MCDI scores – thus allowing to obtain a more robust result than with a leave-one-subject-out scheme. We then gathered all the predicted scores and evaluated their similarity to the true scores by both performing a linear regression and computing the mean absolute error. The statistical significance of this evaluation was assessed in a non-parametric way using a permutation test<sup>82</sup>, which allows estimating the empirical distribution of the computed scores under the null hypothesis of no effect. This statistical procedure was necessary to account for the non-independence of the predicted scores, induced by the cross validation.

## References

- Partanen, E. *et al.* Learning-induced neural plasticity of speech processing before birth. *Proc. Natl. Acad. Sci. USA* **110**(37), 15145–15150 (2013).
- Benavides-Varela, S., Hochmann, J. R., Macagno, F., Nespor, M. & Mehler, J. Newborn's brain activity signals the origin of word memories. *Proc. Natl. Acad. Sci. USA* **109**(44), 17908–17913 (2012).
- Lecanuet, J. P., Granier-Deferre, C., Jacquet, A. Y. & DeCasper, A. J. Fetal discrimination of low-pitched musical notes. *Dev. Psychobiol.* **36**, 29–39 (2000).
- Partanen, E., Kujala, T., Tervaniemi, M. & Huotilainen, M. Prenatal music exposure induces long-term neural effects. *PLoS ONE* **8**(10), e78946 (2013).
- Rakic, P. Specification of cerebral cortical areas. *Science* **241**(4862), 170–176 (1988).
- Rakic, P. A small step for the cell, a giant leap for mankind: a hypothesis of neocortical expansion during evolution. *Trends Neurosci.* **18**(9), 383–388 (1995).
- Dehaene-Lambertz, G., Dehaene, S. & Hertz-Pannier, L. Functional neuroimaging of speech perception in infants. *Science* **298**, 2013–2015 (2002).
- Dehaene-Lambertz, G. *et al.* Functional organization of perisylvian activation during presentation of sentences in preverbal infants. *Proc. Natl. Acad. Sci. USA* **103**(38), 14240–14245 (2006).
- Perani, D. *et al.* Neural language networks at birth. *Proc. Natl. Acad. Sci. USA* **108**(38), 16056–16061 (2011).
- Perani, D. *et al.* Functional specializations for music processing in the human newborn brain. *Proc. Natl. Acad. Sci. USA* **107**(10), 4758–4763 (2010).
- Telkemeyer, S. *et al.* Sensitivity of newborn auditory cortex to the temporal structure of sounds. *Journal of Neuroscience* **29**(47), 14726–14733 (2009).
- Zatorre, R. J. & Belin, P. Spectral and temporal processing in human auditory cortex. *Cerebral Cortex* **11**(10), 946–953 (2001).
- Giraud, A. L. *et al.* Endogenous cortical rhythms determine cerebral specialization for speech perception and production. *Neuron* **56**(6), 1127–1134 (2007).
- Winkler, I., Háden, G. P., Ladinig, O., Sziller, I. & Honing, H. Newborn infants detect the beat in music. *Proc. Natl. Acad. Sci. USA* **106**(7), 2468–2471 (2009).
- Nazzi, T., Floccia, C. & Bertoncini, J. Discrimination of pitch contours by neonates. *Infant Behav. and Dev.* **21**(4), 779–784 (1998).
- Granier-Deferre, C., Ribeiro, A., Jacquet, A. Y. & Bassereau, S. Near-term fetuses process temporal features of speech. *Dev. Sci.* **14**(2), 336–352 (2011).
- Mampe, B., Friederici, A. D., Christophe, A. & Wermke, K. Newborns' cry melody is shaped by their native language. *Curr. Biol.* **19**(23), 1994–1997 (2009).
- Shukla, M., White, K. S. & Aslin, R. N. Prosody guides the rapid mapping of auditory word forms onto visual objects in 6-month-old infants. *Proc. Natl. Acad. Sci. USA* **108**(15), 6038–6043 (2011).
- Trainor, L. J. & Desjardins, R. N. Pitch characteristics of infant-directed speech affect infants' ability to discriminate vowels. *Psychonomic Bulletin and Review* **9**(2), 335–340 (2002).

20. Thiessen, E. D., Hill, E. & Saffran, J. R. Infant-directed speech facilitates word segmentation. *Infancy* **7**, 49–67 (2005).
21. Saffran, J. R., Aslin, R. N. & Newport, E. L. Statistical learning by 8-month-old infants. *Science* **274**, 1926–1928 (1996).
22. Thiessen, E. D. & Saffran, J. R. When cues collide: Use of stress and statistical cues to word boundaries by 7- to 9-month-old infants. *Dev. Psychol.* **39**, 706–716 (2003).
23. Bulf, H., Johnson, S. P. & Valenza, E. Visual statistical learning in the newborn infant. *Cognition* **121**(1), 127–132 (2011).
24. Kudo, N., Nonaka, Y., Mizuno, N., Mizuno, K. & Okanoya, K. On-line statistical segmentation of a non-speech auditory stream in neonates as demonstrated by event-related brain potentials. *Dev. Sci.* **14**(5), 1100–1106 (2011).
25. Teinonen, T., Fellman, V., Näätänen, R., Alku, P. & Huotilainen, M. Statistical language learning in neonates revealed by event-related brain potentials. *BMC Neuroscience* **10**, 21 (2009).
26. Cunillera, T., Toro, J. M., Sebastián-Galles, N. & Rodríguez-Fornells, A. The effects of stress and statistical cues on continuous speech segmentation: an event-related brain potential study. *Brain Research* **1123**(1), 168–178 (2006).
27. De Diego Balaguer, R., Rodríguez-Fornells, A. & Bachoud-Lévi, A. C. Prosodic cues enhance rule learning by changing speech segmentation mechanisms. *Front. Psychol.* **6**, 1478 (2015).
28. Johnson, E. K. & Jusczyk, P. W. Word segmentation by 8-month-olds: When speech cues count more than statistics. *J. Mem. and Lang.* **44**, 548–567 (2001).
29. Bosseler, A. N., Teinonen, T., Tervaniemi, M. & Huotilainen, M. Infant Directed Speech Enhances Statistical Learning in Newborn Infants: An ERP Study. *PLoS ONE* **11**(9), e0162177 (2016).
30. François, C., Cunillera, T., Garcia, E., Laine, M. & Rodríguez-Fornells, A. Neurophysiological evidence for the interplay of speech segmentation and word-referent mapping during novel word learning. *Neuropsychologia* **98**, 56–67 (2017).
31. D D Balaguer, R., Toro, J. M., Rodríguez-Fornells, A. & Bachoud-Lévi, A. C. Different neurophysiological mechanisms underlying word and rule extraction from speech. *PLoS ONE* **2**(11), e1175 (2007).
32. François, C., Jaillet, F., Takerkart, S. & Schön, D. Faster sound stream segmentation in musicians than in nonmusicians. *PLoS ONE* **9**(7), e101340 (2014).
33. Cunillera, T. *et al.* Time course and functional neuroanatomy of speech segmentation in adults. *NeuroImage* **48**(3), 541–553 (2009).
34. Alho, K., Sainio, K., Sajaniemi, N., Reinikainen, K. & Näätänen, R. Event-related brain potential of human newborns to pitch change of an acoustic stimulus. *Electroencephalography and Clinical Neurophysiology/Evoked Potentials Section* **77**(2), 151–155 (1990).
35. Háden, G. P., Németh, R., Török, M. & Winkler, I. Predictive processing of pitch trends in newborn infants. *Brain research* **1626**, 14–20 (2015).
36. Näätänen, R., Jacobsen, T. & Winkler, I. Memory-based or afferent processes in mismatch negativity (MMN): A review of the evidence. *Psychophysiology* **42**(1), 25–32 (2005).
37. Kuhl, P. K. Early language acquisition: cracking the speech code. *Nat. Rev. Neurosci.* **5**(11), 831–843 (2004).
38. Newman, R., Ratner, N. B., Jusczyk, A. M., Jusczyk, P. W. & Dow, K. A. Infants' early ability to segment the conversational speech signal predicts later language development: a retrospective analysis. *Dev. Psychol.* **42**(4), 643–655 (2006).
39. Singh, L., Steven Reznick, J. & Xuehua, L. Infant word segmentation and childhood vocabulary development: a longitudinal analysis. *Dev. Sci.* **15**(4), 482–495 (2012).
40. Junge, C., Kooijman, V., Hagoort, P. & Cutler, A. Rapid recognition at 10 months as a predictor of language development. *Dev. Sci.* **15**(4), 463–473 (2012).
41. Cohen, J., Cohen, P., West, S. G. & Aiken, L. S. Applied multiple regression/correlation analysis for the behavioral sciences. Routledge (2013).
42. Friedrich, M. & Friederici, A. D. Word learning in 6-month-olds: fast encoding-weak retention. *J. Cogn. Neuro.* **23**, 3228–3240 (2011).
43. Rodríguez-Fornells, A., Cunillera, T., Mestres-Missé, A. & De Diego-Balaguer, R. Neurophysiological mechanisms involved in language learning in adults. *Philos. Trans. R. Soc. Lond. B. Biol. Sci.* **364**(1536), 3711–3735 (2009).
44. Perruchet, P. & Vinter, A. The self-organizing consciousness. *Behav. Brain. Sci.* **25**, 297–300 (2002).
45. Stefanics, G. *et al.* Newborn infants process pitch intervals. *Clin. Neurophysiol.* **120**(2), 304–308 (2009).
46. Virtala, P., Huotilainen, M., Partanen, E., Fellman, V. & Tervaniemi, M. Newborn infants' auditory system is sensitive to Western music chord categories. *Front. Psychol.* **4**, 492 (2013).
47. Dehaene-Lambertz, G. Cerebral specialization for speech and non-speech stimuli in infants. *J. Cogn. Neurosci.* **12**(3), 449–460 (2000).
48. Leppänen, P. H. *et al.* Newborn brain event-related potentials revealing atypical processing of sound frequency and the subsequent association with later literacy skills in children with familial dyslexia. *Cortex* **46**(10), 1362–1376 (2010).
49. Guttorm, T. K. *et al.* Brain event-related potentials (ERPs) measured at birth predict later language development in children with and without familial risk for dyslexia. *Cortex* **41**(3), 291–303 (2005).
50. Trehub, S. E., Trainor, L. J. & Unyk, A. M. Music and speech processing in the first year of life. In H. W. Reese (Ed.), *Advances in child development and behavior* (Vol. 24, pp. 1–35). New York: Academic Press (1993).
51. Fernald, A. Maternal vocalisations to infants as biologically relevant signals: An evolutionary perspective. In Barkow, J. H. Cosmides, L. & Tooby, J. (Eds), *The adapted mind: Evolutionary psychology and the generation of culture*. Oxford: Oxford University Press (1992).
52. Schön, D. *et al.* Song as an aid for language acquisition. *Cognition* **106**(2), 975–983 (2008).
53. Christiansen, M. H., Allen, J. & Seidenberg, M. S. Learning to segment speech using multiple cues: a connectionist model. *Lang. Cogn. Process.* **13**, 221–268 (1998).
54. Fernald, A. Four-month-old infants prefer to listen to motherese. *Infant Behav. and Dev.* **8**, 181–195 (1985).
55. Fernald, A. Approval and disapproval: Infant responsiveness to vocal affect in familiar and unfamiliar languages. *Child Dev.* **64**, 657–667 (1993).
56. Trainor, L. J., Austin, C. M. & Desjardins, R. N. Is infant-directed speech prosody a result of the vocal expression of emotion? *Psychol. Sci.* **11**(3), 188–195 (2000).
57. Blood, A. J. & Zatorre, R. J. Intensely pleasurable responses to music correlate with activity in brain regions implicated in reward and emotion. *Proc. Natl. Acad. Sci. USA* **98**(20), 11818–11823 (2001).
58. Koelsch, S. Towards a neural basis of music-evoked emotions. *Trends Cogn. Sci.* **14**(3), 131–137 (2010).
59. Craik, F. I., Govoni, R., Naveh-Benjamin, M. & Anderson, N. D. The effects of divided attention on encoding and retrieval processes in human memory. *J. Exp. Psychol. Gen.* **125**(2), 159–180 (1996).
60. Adcock, R. A., Thangavel, A., Whitfield-Gabrieli, S., Knutson, B. & Gabrieli, J. D. Reward-motivated learning: mesolimbic activation precedes memory formation. *Neuron* **50**(3), 507–517 (2006).
61. Kaplan, P. S., Bachorowski, J. A., Smoski, M. J. & Hudenko, W. J. Infants of depressed mothers, although competent learners, fail to learn in response to their own mothers' infant-directed speech. *Psychol. Sci.* **13**(3), 268–271 (2002).
62. Ripollés, P. *et al.* The role of reward in word learning and its implications for language acquisition. *Curr. Biol.* **24**(21), 2606–2611 (2014).
63. Kraus, N. & White-Schwoch, T. Unraveling the biology of auditory learning: A Cognitive-Sensorimotor-Reward framework. *Trends Cogn. Sci.* **19**(11), 642–654 (2015).
64. Ylinen, S., Bosseler, A., Junntila, K., Huotilainen, M. Predictive coding accelerates word recognition and learning in the early stages of language development. *Dev. Sci.* <https://doi.org/10.1111/desc.12472> (2016).

65. François, C. & Schön, D. Musical expertise boosts implicit learning of both musical and linguistic structures. *Cerebral Cortex* **21**(10), 2357–2365 (2011).
66. François, C., Chobert, J., Besson, M. & Schön, D. Music training for the development of speech segmentation. *Cerebral Cortex* **23**(9), 2038–2043 (2013).
67. Yao, D. Is the Surface Potential Integral of a Dipole in a Volume Conductor Always Zero? A Cloud Over the Average Reference of EEG and ERP. *Brain Topography* **30**, 161–171 (2017).
68. Liu, Q. *et al.* Estimating a neutral reference for electroencephalographic recordings: the importance of using a high-density montage and a realistic head model. *J. neural eng.* **12**(5), 056012 (2015).
69. Ferjan Ramírez, N., Ramírez, R. R., Clarke, M., Taulu, S., Kuhl, P. K. Speech discrimination in 11-month-old bilingual and monolingual infants: a magnetoencephalography study. *Dev. Sci.* **20**(1) (2017).
70. Bosch, L. & Sebastián-Gallés, N. Native-language recognition abilities in four-month-old infants from monolingual and bilingual environments. *Cognition* **65**(1), 33–69 (1997).
71. Bosch, L. & Sebastián-Gallés, N. Evidence of early language discrimination abilities in infants from bilingual environments. *Infancy* **2**(1), 29–49 (2001).
72. Byers-Heinlein, K., Burns, T. C. & Werker, J. F. The roots of bilingualism in newborns. *Psychological science* **21**(3), 343–348 (2010).
73. Franco, A., Cleeremans, A. & Destrebecqz, A. Statistical learning of two artificial languages presented successively: how conscious? *Front. Psychol.* **2**, 229 (2011).
74. Zhao, T. C. & Kuhl, P. K. Musical intervention enhances infants' neural processing of temporal structure in music and speech. *Proc. Natl. Acad. Sci. USA* **113**(19), 5212–5217 (2016).
75. López-Ornat, S. *et al.* Inventario de Desarrollo Comunicativo MacArthur, adaptación española. Madrid: TEA Ediciones, S.A. (2005).
76. Bayley, N. Bayley Scales of Infant and Toddler Development. 3rd edn. San Antonio, TX: Harcourt Assessment Inc (2006).
77. Partanen, E., Pakarinen, S., Kujala, T. & Huotilainen, M. Infants' brain responses for speech sound changes in fast multifeature MMN paradigm. *Clin. Neurophysiol.* **124**(8), 1578–1585 (2013).
78. Leipälä, J. A., Partanen, E., Kushnerenko, E., Huotilainen, M. & Fellman, V. Perinatal cerebral insults alter auditory event-related potentials. *Early human development* **87**(2), 89–95 (2011).
79. Mahmoudzadeh, M., Wallois, F., Kongolo, G., Goudjil, S. & Dehaene-Lambertz, G. Functional maps at the onset of auditory inputs in very early preterm human neonates. *Cerebral Cortex* **27**(4), 2500–2512 (2017).
80. Gabrieli, J. D., Ghosh, S. S. & Whitfield-Gabrieli, S. Prediction as a humanitarian and pragmatic contribution from human cognitive neuroscience. *Neuron* **85**(1), 11–26 (2015).
81. Varoquaux, G. *et al.* Assessing and tuning brain decoders: Cross-validation, caveats, and guidelines. *NeuroImage* **145**(Pt B), 166–179 (2017).
82. Anderson, M. J. & Robinson, J. *Permutation Tests for Linear Models.* *Australian & New Zealand Journal of Statistics* **43**(1), 75–88 (2001).

## Acknowledgements

We thank all the families and their babies for participating in the study. We also thank the medical staff from the maternal care unit in Sant Joan de Déu Hospital. We thank R. de Diego Balaguer, T. Cunillera for their constructive comments on the manuscript; and D. Cucurell and J. Solé for their help in different steps of the study. This research has been supported by a grant from FYSSEN foundation (post-doctoral grant) as well as a Spanish MINECO project (PSI2015-69132P) and the Catalan government (PERIS2017) to CF, and Spanish MINECO projects (PSI2011-25376 and PSI2014-55105P) to LB.

## Author Contributions

C.F., L.B. and A.R.F. designed the research, C.F. and M.T. performed the research, C.F. and S.T. analysed the data, T.A. provided access to the participants, C.F., L.B. and A.R.F. wrote the manuscript.

## Additional Information

**Supplementary information** accompanies this paper at <https://doi.org/10.1038/s41598-017-12798-2>.

**Competing Interests:** The authors declare that they have no competing interests.

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



**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2017