INFANCY

THE OFFICIAL JOURNAL OF THE INTERNATIONAL CONGRESS OF INFANT STUDIES

Infancy, ***23***(6), 770–794, 2018
© International Congress of Infant Studies (ICIS)
ISSN: 1525-0008 print / 1532-7078 online
DOI: 10.1111/infa.12256

WILEY Blackwell

Individual Differences in Infant Speech Segmentation: Achieving the Lexical Shift

Evan Kidd
The Australian National University and
ARC Centre of Excellence for the Dynamics of Language and
Max Planck Institute for Psycholinguistics

Caroline Junge
Utrecht University

Tara Spokes and Lauren Morrison
The Australian National University and
ARC Centre of Excellence for the Dynamics of Language

Anne Cutler
ARC Centre of Excellence for the Dynamics of Language and
Max Planck Institute for Psycholinguistics and
Western Sydney University

We report a large-scale electrophysiological study of infant speech segmentation, in which over 100 English-acquiring 9-month-olds were exposed to unfamiliar bisyllabic words embedded in sentences (e.g., *He saw a wild eagle up there*), after which their brain responses to either the just-familiarized word (*eagle*) or a control word (*coral*) were recorded. When initial exposure occurs in continuous speech, as here, past studies have reported that even somewhat older infants do not reliably recognize target words, but that successful segmentation varies across children. Here, we both confirm and further uncover the nature of this variation. The segmentation response systematically varied across individuals and was related to their vocabulary development. About one-third of the group showed a left-frontally located relative negativity in response to familiar versus control targets, which has previously been described as a mature response. Another third showed a similarly located positive-going reaction (a previously described immature response), and the remaining third formed an intermediate grouping that was primarily characterized by an initial response delay. A fine-grained group-level analysis suggested that a developmental shift to a lexical mode of processing occurs toward the end of the first year, with variation across individual infants in the exact timing of this shift.

Correspondence should be sent to Evan Kidd, Max Planck Institute for Psycholinguistics, PO Box 310, AH Nijmegen 6500, The Netherlands. E-mail: evan.kidd@mpi.nl

People differ in every imaginable way, and their individual differences are visible from the earliest recordable moments of their development (Kidd, Donnelly, & Christiansen, 2018). Inasmuch as individual differences at an earlier developmental stage predict correlated differences at a later stage, there is significant practical and theoretical value in understanding how they pattern and interact with developmental change. Yet studies of individual differences across time are rare; partly because longitudinal research is costly, but also in part because of the vicious circle arising from the rarity: to avoid fishing expeditions, longitudinal studies need a pointer to which of the many dimensions of difference are most likely to reward investigation.

In language acquisition, one factor known to vary widely across individuals and to predict several other linguistic measures later in development is vocabulary size. Thus, vocabulary size predicts not only subsequent vocabulary growth but also morphosyntactic development (e.g., Fenson et al., 1994; Hoff, Quinn, & Giguere, 2017; Marchman & Bates, 1994). Charting this dependence is a critically important area of acquisition research, as it provides insight into the relatedness of different linguistic skills. This knowledge has both theoretical and practical importance. In theoretical terms, identifying relatedness across domains informs and constrains theory (Bates, Dale, & Thal, 1995). In practical terms, identifying predictors of subsequent development directs practitioners to early domains of language that will particularly inform diagnosis and treatment of language delay.

But although words are early targets of talking, expressive vocabulary size measures cannot tap into the initial stages of native-language learning. Such vocabulary assessments are usually undertaken from the second year of life on. In the first year of a child's life, however, a wide range of language knowledge is being laid down, while the process of building a vocabulary is just beginning (Johnson, 2016; Saffran, Werker, & Werner, 2006). In that initial year, children's language acquisition work is largely perceptual. If early perceptual performance proves to display significant individual differences, and these differences in turn prove to correlate with later vocabulary size, then the chain of prediction can be usefully extended—from the precursors of word learning in the first year, to sentence- and discourse-level skills years later.

There are encouraging signs that early speech perception indeed relates to other language skills (Cristia, Seidl, Junge, Soderstrom, & Hagoort, 2014). For instance, Tsao, Liu, and Kuhl (2004) compared the perceptual discrimination scores of 28 6-month-olds for a pair of non-native vowels against parental reports of their language skills 7, 10, and 18 months later. The better the discrimination scores had been at 6 months, the more words the children knew and/or produced later on.

Brain responses to speech also show a relation to later language skills. Molfese and Molfese (1985, 1997) found that groups of children later classified as performing better or worse than average on language tasks differed in the pattern of their auditory evoked potentials to consonant-vowel syllables that had been recorded at birth. Rivera-Gaxiola, Klarman, Garcia-Sierra, and Kuhl (2005) found that 11-month-olds with a negative-going response to a deviant phoneme among repeating standards had higher vocabulary over the following year and a half than age-mates with a positive-going response in the same task.

Particularly promising, however, are relations that appear in a task known to form part of the process of establishing an initial store of potential word forms: speech segmentation. The input that infants hear consists largely of continuous speech (Aslin, Woodward, LaMendola, & Bever, 1996; van de Weijer, 1998); to store potential word

forms in a nascent vocabulary, infants must extract them from the running-speech context. Unsurprisingly, this turns out to be easier for words that are at utterance edges (Seidl & Johnson, 2006); but infants actually begin by 7 to 9 months of age to show segmentation of words bounded on both sides by speech. This was first established by Jusczyk and Aslin (1995), who devised a version of the Headturn Preference Paradigm (HPP) in which infants were first familiarized with words and then heard short spoken texts containing (or not containing) those words. Segmentation of the familiarized forms from their contexts was deemed to have occurred if the infants showed a preference for the speech that contained those forms.

Performance in this task correlates with later vocabulary mastery. For instance, Newman, Bernstein Ratner, Jusczyk, Jusczyk, and Dow (2006) tested the linguistic skills of children who had participated as infants in HPP segmentation studies. Two-year-olds in the top 15% for vocabulary skills were significantly more likely to have successfully segmented speech as infants than age-mates in the bottom 15%. A subset of these children were followed until age 4–6 years, with the language differences at 2 years proving still significant at the later time point. Similar relations of HPP segmentation scores to later skills have been reported by Newman, Rowe, and Bernstein Ratner (2016) and Singh, Reznick, and Xuehua (2012). Note that these results are suggestive: Segmentation ability was inferred when infants preferred listening to familiarized words over control words. Due to its reliance on infant preference, the HPP paradigm may not be the most sensitive measure of individual differences.

More sensitive individual-level measurement of speech segmentation was enabled when Kooijman, Hagoort, and Cutler (2005) adapted Jusczyk and Aslin's experiment as a brain response paradigm. Kooijman et al. recorded event-related potentials (ERPs) in the infants' brains as they listened to blocks of target words in isolation (familiarization phase) followed by sentences (test phase). The target (familiarized) words occurred in half of the sentences; the other sentences contained a matched but unfamiliar word. A difference in the ERPs to familiarized versus unfamiliar words indicated that recognition of the familiarized words had occurred, that is, that the target words had been segmented from the surrounding speech. Some children could successfully segment even if both the first exposure and later test involved continuous speech (Junge, Cutler, & Hagoort, 2014).

The recognition response observed in these studies was a left-lateralized frontal negative-going waveform in response to a familiarized word, compared to an unfamiliarized baseline word. This same response was subsequently observed in infants at equivalent developmental stages for French (Goyet, de Schonen, & Nazzi, 2010) and German (Männel & Friederici, 2013). The negativity is similar to negative-going responses observed in older infants responding to known versus unknown words heard in isolation (e.g., Mills, Coffey-Corina, & Neville, 1993, 1997; Thierry, Vihman, & Roberts, 2003). This latter negativity increases in amplitude with development (Männel & Friederici, 2013; Vihman, Thierry, Lum, Keren-Portnoy, & Martin, 2007), and often co-occurs with a later negativity around 600–900 msec from word onset (Mills et al., 1993, 1997).

Importantly, the segmentation findings with ERPs also correlate with later language abilities, across several samples (Junge & Cutler, 2014; Junge, Kooijman, Hagoort, & Cutler, 2012; Kooijman, Junge, Johnson, Hagoort, & Cutler, 2013). Kooijman, Junge et al. tested Dutch-acquiring 7-month-olds, who first heard ten successive presentations of isolated bisyllabic words (e.g., *monnik*, "monk"), then sentences with or without the

familiarized word. Unlike the 10-month-olds in Kooijman et al. (2005), who typically showed the left frontal negativity in this situation, these younger infants fell into two groups: a smaller group who did show this negative response, and a somewhat larger group who instead showed a response on the same left front electrodes that was *positive* relative to the baseline. The two groups were followed up when they were, on average, 3 years old (range: 28–47 months). At follow-up, the negative responders had significantly higher language comprehension and word production skills in comparison with the positive responders. Furthermore, the size of the amplitude difference between familiarized and unfamiliarized words over the left frontal electrodes correlated significantly with word production ($r^2 = .22$, $p < .05$).

Junge, Kooijman, et al. (2012) found that segmentation responses in the brain at 10 months likewise correlated with later language skills. In their study, infants were familiarized with a target word in a single sentence (e.g., *Die kleine mammoet zwemt in de rivier*, "That small mammoth swims in the river") and then were played either the target word (*mammoet*) or a control word (e.g., *hommel*, "bumblebee") 300 msec later. This is a harder task than hearing familiarization tokens in isolation, and, averaged over the whole group, 10-month-olds could not successfully perform it (i.e., recognize the words at test). However, when the children were later divided into groups by vocabulary size at 12 and at 24 months, the high-vocabulary group showed a broadly distributed negativity in response to the target word 200–650 msec postword onset, whereas the low-vocabulary group showed little evidence of any recognition response. Once again, the ERP amplitude measures were informative: There were significant correlations between the familiar-control difference in amplitude at 10 months and CDI scores of word and sentence comprehension at 12 months ($r^2 = .35$ & $r^2 = .30$, respectively, p's $< .01$), and between the amplitude measure and word comprehension and production at 24 months ($r^2 = .26$, $p < .01$, & $r^2 = .15$, $p < .05$, respectively).

Junge et al. (2014) exposed 10-month-olds to words in continuous speech during both familiarization and test, making the task yet harder. In familiarization, eight short sentences (like those in their earlier studies) were presented, all containing the target word. At test, there were four sentences, two containing the target and two containing control words. Most of their participants ($n = 19$) showed a broadly distributed negativity during the test phase and indeed showed this effect during familiarization, across the repetitions. Some ($n = 9$), however, displayed a positive-going response at test and no sign of increasing familiarity during familiarization. Junge and Cutler (2014) then followed these children up at 16 months, at which time their eye movements were recorded to pictures (e.g., a ball, a cat) as they heard sentences (e.g., "Look at the ball!"). In this lexical processing task, the negative responders had more robust lexical representations, as indicated by a greater tendency to fixate on target pictures, in comparison with positive responders.

Thus, the presence of a speech segmentation effect in the infant brain (a negative-going rather than a positive-going response) varies significantly across groups of infants, and where a response occurs, its relative amplitude varies too. Both measures, moreover, have been reported to be predictive of language development in the second and third year of life. These past ERP studies have had relatively small sample sizes and have only reported simple correlations of measures of segmentation with later language. Although meta-analyses suggest these correlations are stable (Cristia et al., 2014; Junge & Cutler, 2014), there are several reasons why further verification data are called for. Firstly, small sample sizes can inflate the size of correlations and, in general,

can provide poor estimates of the true nature of an effect in the population (Oakes, 2017); therefore the true strength of the predictive relationship between segmentation skill and later language development may have been overestimated. Secondly, we do not know the extent to which ERP measures of segmentation are reliable (for a related discussion of behavioral measures see Cristia, Seidl, Singh, & Houston, 2016). Given the recent importance placed on replicability in infancy research and beyond (Frank et al., 2017; Open Science Collaboration, 2015), an important next step is to determine whether these results replicate in a larger sample of infants than has been previously tested.

To this end, we report an ERP study of segmentation with more than 100 9-month-old participants, in which we use the relatively difficult measure involving familiarization with continuous speech followed by isolated words at test (Junge, Kooijman, et al., 2012). In keeping with the evidence from the ERP segmentation studies reviewed above, we expect that a subset of our participant group will show the left frontal negative-going recognition response regarded as the mature response, while another subset will display a positive-going response, which, both for speech segmentation (Junge & Cutler, 2014; Kooijman et al., 2013) and for phonemic processing (Rivera-Gaxiola, Klarman, et al., 2005), is regarded as an immature response. Viewing the progress from immature to mature response as a continuous development from positive to negative, we further expect that with a sufficiently large group the full trajectory of this development will be more completely revealed. That is, we expect there to be participants showing a response that is developmentally intermediate between the immature and mature stages. Follow-up assessment of the children's language development at 12 and 15 months allows these groupings to be verified and enables fine-grained analyses of each group to better categorize the psychophysiological processes underlying word processing in the first year.

METHOD

Participants

A total of 113 9-month-old infants (62 male; 51 female) who were acquiring Australian English as a native language were recruited from a mid-sized city in Australia. This testing session was the first in a four year study designed to investigate language development using multiple research methods. Inclusion criteria were as follows: (1) full-term (at least 37 weeks gestation) babies born with a typical birthweight (>2.5 kg), (2) a predominantly monolingual language environment (no more than 20% exposure to a language other than English), and (3) no history of medical conditions that would affect typical language development, such as repeated ear infections, visual or hearing impairment, or diagnosed developmental disabilities. Consistent with the demographics of the city, the sample was drawn from families high in socioeconomic status. More than 75% of the parents had completed a bachelor degree or higher. Seven boys were excluded because they did not provide enough data across each experimental condition (10 valid trials per condition) to be included in the data analysis. A further three male children were excluded because they were subsequently diagnosed, after 2 years of age, with (i) global developmental delay (including language) (1), or (ii) partial hearing loss (2). The final dataset included 103 infants, with 52 males and 51 females and a mean age of 9.5 months (SD = 6.9 days, range = 276–308 days). Parents were given $10 reimbursement to cover travel expenses.

The study was conducted according to Declaration of Helsinki guidelines, with written informed consent obtained from a parent or guardian for each child before data collection. All procedures were approved by the Human Research Ethics Committee at The Australian National University (protocol no. 2015/007).

Materials

Audio stimuli

A total of 120 disyllabic trochaic words were used as target words across 240 sentences (e.g., *barrel, dative, jester, gamble*, full materials available from authors upon request). The chosen words were all low frequency and did not appear in the MacArthur-Bates CDI Words and Gestures vocabulary inventory (Fenson et al., 2007) and thus would be unlikely to be known by the children. Sentences were 7–8 syllables long and the target words appeared sentence medially (i.e., not in sentence initial or final position). Sentences and target words were recorded by a female speaker of Australian English. The average length of sentences was 3,501 msec. The average target word length in the exposure sentence was 798 msec, and at test, when the word was presented in isolation, was 930 msec.

Following Junge, Kooijman, et al. (2012), the presentation of each sentence was followed by a test word that was either in the preceding sentence (direct familiar condition; DF) or not in it (non-direct familiar condition; nDF). There were 120 DF trials and 120 nDF trials presented across four blocks. Of the nDF trials, 60 had a test word that had not been presented at all before (*unfamiliar*, henceforth UF), while the other 60 had a test word that had been presented as a DF word 10–12 trials prior (*indirect familiar*, henceforth IF). That is, for the IF condition, there had been two presentations of the target word 10–12 trials earlier, once in continuous speech and once in isolation. This follows previous studies (e.g., Goyet et al., 2010; Junge, Kooijman, et al., 2012), where analyses have revealed that infants do not appear to recognize IF targets. However, small sample sizes and low numbers of valid trials across conditions in past studies may have resulted in low power, and so we thought it prudent to analyze the IF and UF conditions separately, particularly in light of the possibility that children with good segmentation skills, and therefore, strong phonemic representations may indeed recognize IF words despite the delay between presentation and target (as do adults, Snijders, Kooijman, Cutler, & Hagoort, 2007). The presentation order of the four blocks was counterbalanced across participants. The audio input was paused when needed (e.g., when a child became restless), and the state of the infant was checked at the end of each block. An audio attention grabber was presented after every 10 trials. Attention grabbers ranged in duration from 1,453 to 5,912 msec (mean length = 3,134 msec) and consisted of non-speech sounds such as bugle calls, a "magic wand" sound, and children's laughter. This was intended to maintain the infant's interest in the ongoing audio (Stets, Burt, & Reid, 2013). Portable desktop speakers were positioned in front and on either side of the screen and approximately 75 cm in front of the infant. The auditory stimuli ranged in volume from 80 to 90 dB.

EEG data acquisition

Continuous EEG was recorded using a 32-channel ActiveTwo BioSemi system (Amsterdam, Netherlands), which was configured to the standard International 10–20

System. The EEG data were recorded relative to common mode sense and driven right leg (CMS/DRL) electrodes placed at the back of the head, to the left and right of a midline parietal-occipital electrode (POz), respectively. Continuous EEG was acquired using Actiview (version 7.05), and the stimuli were presented using E-Prime software (version 2.0). EEG data were recorded with a 2,048 Hz sample rate. Electrode offsets were <20 μV.

Language and cognitive development assessments

Children's language and cognitive development were assessed via parental questionnaire. Language was measured using the MacArthur-Bates CDI—Words and Gestures (Fenson et al., 2007). Cognitive development was measured using The Ages and Stages Questionnaire (ASQ, Squires & Bricker, 2009). These were given to the parents at the end of the testing session; parents completed them at home and returned them to the researchers via mail. Parents completed these questionnaires again when their child was 12 and 15 months of age.

Procedure

Although intended for MEG, the reporting guidelines of Gross et al. (2013) were applied where appropriate for EEG. These included: (1) the preparation of the system prior to the participant's arrival, (2) the design of the experiment, and (3) the preparation of the participant. In the infant population, this included requesting that parents bring substantial food for the infant. The cap was pre-gelled and fitted, and the infant was encouraged to eat until satisfied. Pre-gelling minimizes setup time, as children do not have to sit through the process. Allowing the child to eat or drink allows the gel time to warm, providing a better signal, and also means the infant is likely to be more settled and sustain their attention for longer. This is important, as infants are notoriously fussy participants in EEG research. Next, the infant was placed in an infant car seat in the experimental area and data collection initiated. The car seat was used to minimize movement. A non-synchronized stream of age-relevant cartoons was presented on a 22″ monitor positioned approximately 90 cm in front of the infant. One researcher sat with the infant to maintain engagement and minimize head movement, using a range of silent toys and puppets. The parent was invited to sit in the room but directly behind the infant so that they could not be seen by the infant. In our experience, infants are more likely to stay positioned when they cannot hear or see their parent. A second researcher ran the EEG acquisition from behind a freestanding wall partition.

Pre-processing & statistical analyses

Continuous EEG data were processed offline with the BESA software (version 5.3.6, GmbH, Gräfelfing, Germany) using an average reference. Epochs were defined for a period from 300 msec prior to the onset of the test word to 1,000 msec post-onset. Baseline definition was taken from 200 msec prior to word onset to word onset. Following referencing, filters were applied from 300 msec to allow for early effects of the filters to pass from 300 to 200 msec prestimulus. A bandpass filter of 0.1 Hz (6 db/oct; forward) to 30 Hz (24 db/oct; zero phase) was applied prior to artifact rejection.

EEG epochs were rejected when changes in amplitude exceeded 200 μV. Time windows for mean amplitude analysis were considered by visual inspection of the grand average waveform.

Event-related potentials were compared between conditions using a series of ANOVAs to compare ERP mean amplitudes at two time windows (250–500 and 600–800 msec postonset of the test word). In previous studies, the electrode array was divided into four quadrants (e.g., Junge, Kooijman, et al., 2012; Kooijman et al., 2005), with results from the left frontal quadrant used to distinguish positive from negative responders (Junge & Cutler, 2014; Junge, Kooijman, et al., 2012), and for comparability with the prior findings, we grouped the lateral electrodes in the same manner: Left Frontal (AF3, F3, F7, FC5), Right Frontal (AF4, F4, F8, FC6), Left Parietal (P7, P3, P03, O1), Right Parietal (P8, P4, PO4, O2). Significant effects were pursued with post hoc analyses. Non-integer degrees of freedom were Greenhouse-Geisser corrected for violation of sphericity. Note that although the signal-to-noise ratios are unlikely to be equal across conditions (i.e., the ERP for DF is based on twice as many trials as the ERP for UF), this should not be problematic for our dependent variables that report area under the curve (mean voltage over a time window); such values are less susceptible to differences in trial inclusions (Luck, 2010).

RESULTS

The mean number of accepted trials was 54.71 (SD = 19.25) for the DF condition; 27.38 (SD = 9.79) for the IF and 28.76 (SD = 9.50) for the UF conditions. Figure 1 plots the ERPs for the corresponding conditions for each quadrant. The frontal quadrants show an early positivity (250–500 msec), with an approximate peak of 2 μV, followed by a late negativity at 600–800 msec. The posterior quadrants show an extended negativity of approximately -4 μV, which starts somewhat earlier (at 150 msec) than the early frontal positivity. Past research has found evidence for segmentation in frontal areas. We thus largely limit our analyses to the early 250–500 msec time window (as per Junge, Kooijman, et al., 2012).

A 3 (condition: DF, IF, UF) $\times$ 4 (quadrant: Left Frontal, Right Frontal, Left Parietal, Right Parietal) way repeated-measures ANOVA for the 250–500 msec time window suggested that there was no significant difference between the three conditions across the four quadrants [$F(2, 204) = .445$, $p = .64$, $\eta_p^2 = .004$] and no condition x quadrant interaction [$F(6, 612) = 1.13$, $p = .35$, $\eta_p^2 = .011$]. We also checked whether there were any differences during the 600–800 msec time window. Here, we found a significant condition x quadrant interaction [$F(3.74, 381.34) = 3.98$, $p = .004$, $\eta_p^2 = .038$]. Post hoc analyses within each region revealed a slight trend toward a main effect of condition in the left frontal region [$F(2, 204) = 2.37$, $p = .096$, $\eta_p^2 = .023$], which appeared to drive the interaction. Post hoc Bonferroni comparisons revealed that in this region the DF condition was more negative than the UF ($p = .046$). There were no other significant effects of condition in the other quadrants [Fs < 1.85, ps $> .15$].

Individual differences analyses

At the group level, we observed some weak evidence for segmentation in the late 600–800 msec time window, but not in the earlier time window 250–500, which usually is

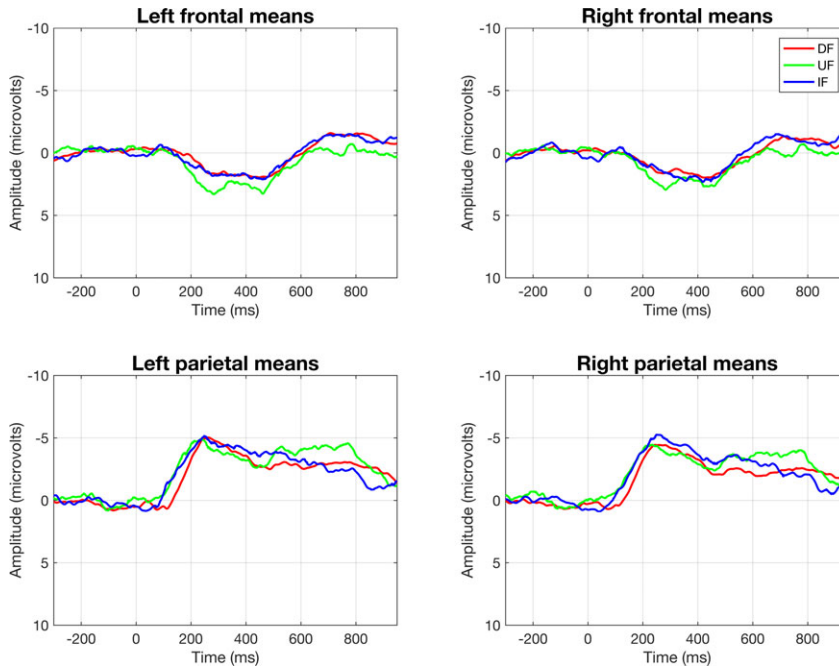

Figure 1 Grand average event-related potential of direct familiar (red line), indirect familiar (blue line), and unfamiliar (green line) conditions at the Left Frontal (LF), Right Frontal (RF), Left Parietal (LP), and Right Parietal (RP) regions for all participants (n = 103). Negativity is plotted upwards.

the key window of interest tapping word recognition (Mills, Conboy, & Paton, 2005). Junge, Kooijman, et al. (2012) further analyzed their data by adding a vocabulary group (high versus low) as a between-participants variable (where vocabulary was measured at 12 months). We replicated this analysis over our 250–500 msec time window, performing a median split on vocabulary comprehension at 12 months to create high- and low-vocabulary groups. This analysis included 99 children, as four children had either withdrawn from the study or their parents had failed to return their CDI at this session. Consistent with Junge, Kooijman, et al., there was a significant condition x quadrant x vocabulary group interaction [$F(6, 582) = 2.89$, $p = .009$, $\eta_p^2 = .029$]. Follow-up analyses for the low-vocabulary group revealed no effect of condition and no condition x quadrant interaction (Fs < 1). However, in the high-vocabulary group there was a significant condition x quadrant interaction [$F(4.54, 222.2) = 3.53$, $p = .006$, $\eta_p^2 = .067$]. Follow-up analyses within each region revealed significant main effects of condition in both the left frontal [$F(2, 98) = 3.67$, $p = .029$, $\eta_p^2 = .07$] and right frontal quadrants [$F(2, 98) = 3.42$, $p = .037$, $\eta_p^2 = .065$]. Post hoc Bonferroni comparisons revealed that, for the left frontal quadrant, the DF condition was significantly more negative than the UF ($p = .01$), and in the right frontal quadrant, the DF condition was significantly more negative than the IF ($p = .048$). There were no other significant comparisons.

Consistent with Junge, Kooijman, et al.'s (2012) Dutch-speaking children, we have evidence of segmentation of target words from one exposure sentence in English-

speaking children who have comparatively good vocabulary skills for their age. In children with higher vocabulary, we found the more mature relative negativity to familiar words, and in the children with low vocabulary, there was no difference in ERP amplitude across the conditions.[1] Junge, Kooijman, et al. also reported significant correlations between the average difference to familiarized (i.e., DF) and unfamiliarized words (i.e., nDF) in the left frontal region and children's subsequent language development. We found a similar though weaker pattern of association: The average difference score was negatively correlated with language comprehension, as measured by the CDI, at 12 and 15 months (12 months: Spearman's $\rho = -.20$, $p = .048$, two-tailed, $N = 99$; 15 months: Spearman's $\rho = -.29$, $p = .005$, two-tailed, $N = 92$; nonparametric correlations were used because the vocabulary data were not normally distributed). Note that the negative correlation indicates that the amplitude of the negative ERP difference wave increases with larger vocabulary scores. Also consistent with this study is that there was no association with language production at these ages, presumably because at this stage, most infants produce only a few words, thus leading to a floor effect. We conducted additional regression analyses to determine whether the association between the ERP signature of segmentation and subsequent language development held after controlling for the additional variables of: (1) language comprehension at 9 months, (2) gender, (3) SES (as measured by combined educational attainment of primary and secondary caregivers), (4) age (in days) at EEG experimental session. The distribution of the children's vocabulary scores was not normal and contained several influential cases that weakened any multiple linear regression analysis (see Appendix S1). We therefore report the results of two logistic regression analyses, where the dependent measure is binary (0 = low vocabulary, 1 = high vocabulary, with groups created via a median split). The four control variables were entered in the first block of the regression, followed by the left frontal ERP difference score in the second block. Therefore, we were testing whether the addition of children's segmentation skills explained any additional variance in children's subsequent vocabulary knowledge at 12 and 15 months over and above our control variables. In both cases, at Block 1 the models were significant (12 month model: $\chi^2 = 59.94$, $df = 4$, $p < .001$, Nagelkerke $R^2 = .615$; 15 month model, $\chi^2 = 42.55$, $df = 4$, $p < .001$, Nagelkerke $R^2 = .498$), with only vocabulary comprehension at 9 months significantly contributing to the model. The addition of our segmentation measure at Block 2 significantly added to the models (12 month model, $\chi^2 = 5.78$, $df = 1$, $p = .016$, Final 12 month model: $\chi^2 = 65.71$, $df = 1$, $p < .001$, Nagelkerke $R^2 = .656$; 15 month model, $\chi^2 = 9.85$, $df = 1$, $p = .002$, Final 15 month model: $\chi^2 = 52.40$, $df = 5$, $p < .001$, Nagelkerke $R^2 = .584$). Table 1 shows the results for the final models.

Interim summary

Our results replicate and extend those of Junge, Kooijman et al. (2012). Using a demanding segmentation task where children were only exposed to one instance of

[1]An anonymous reviewer suggested that the relatively high SES of our sample might pose problems for our median split of the data because the children's vocabulary might be higher compared to the general population. We therefore compared our sample's vocabulary at 12 and 15 months to the Australian English norms (Kalashnikova, Schwarz, & Burnham, 2016), available from WordBank (Frank, Braginsky, Yurovsky, & Marchman, 2017). In both cases, the distribution of vocabulary scores in our sample was statistically indistinguishable from the norming sample (12 months: $Z = -.33$, $p = .74$; 15 months: $Z = -.33$, $p = .97$).

TABLE 1
Results of Logistic Regression Predicting Vocabulary Comprehension at 12 and 15 Months

	β	$SE\ (\beta)$	e^{β}	*Wald*	*df*	*p*
12 months[a]						
Constant	10.33	14.73	30,535.86	.49	1	.48
Vocabulary (9 months)	.12	.03	1.13	17.49	1	<.001***
Age	−.044	.05	.96	.75	1	.39
Gender	.54	.62	1.72	.76	1	.38
SES	−.11	.16	.90	.48	1	.49
Segmentation	−.13	.06	.88	4.91	1	.027*
15 months[b]						
Constant	−6.56	13.74	.001	.23	1	.63
Vocabulary (9 months)	.084	.02	1.09	16.37	1	<.001***
Age	.011	.047	1.01	.059	1	.81
Gender	1.03	.60	2.97	2.97	1	.085
SES	.045	.14	1.05	.11	1	.75
Segmentation	−.16	.056	.85	8.32	1	.004**

[a]$N = 97$, [b]$N = 91$.
***$p < .001$, **$p < .01$, *$p < .05$.

the target word in a continuous-speech context, we found that only children who later had higher vocabularies differentiated between target and non-target words at test, showing a mature ERP response. We also found that the magnitude of the difference in ERP response to target and non-target words was longitudinally associated with vocabulary size three and six months later, a result that held after controlling for several factors that could have potentially contributed to the previously reported associations but were not controlled, such as concurrent vocabulary at time of testing.

Our results also suggest significant individual variability in segmentation skills in English-speaking infants at 9 months. The two types of ERP response isolated in prior work—a mature relative negativity and an immature relative positivity—are seen in the negative relation we found between our segmentation measure and linguistic maturity. In the next set of analyses, we probe further into the distribution and developmental trajectory of the ERP response.

Tracking the developmental trajectory of the segmentation response

There is systematic variation in the ERP response in the left frontal region. The average standard deviation (5.38 μV) across the three conditions is double the average amplitude (2.68 μV). Figure 2 shows the density plot of the difference between the DF and UF conditions (our putative baseline). Despite three extreme values contributing to a slight positive skew, the distribution of DF–UF difference scores is normally distributed (Shaprio–Wilks test, $p = .166$). In line with past findings, we have both clear negative and positive responder subgroups, but as hypothesized, the normal distribution suggests an intermediate subgroup between the clearly negative and clearly positive. We next analyzed the ERPs of these children to better understand the range of segmentation skills in the sample. The sample was divided

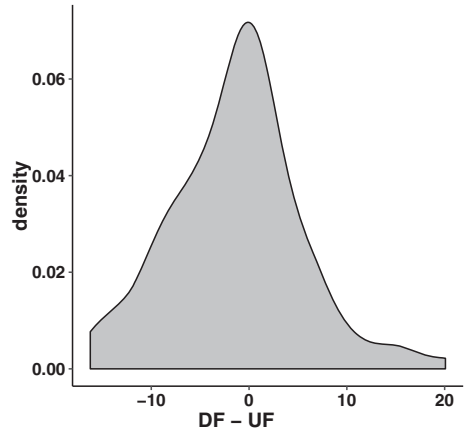

Figure 2 Density plot of average event-related potential difference scores between direct familiar and unfamiliar trials (in μV).

into three subgroups according to their average difference between the DF and UF conditions, via a ternary split. Those children who had a negative-going mean amplitude difference exceeding -3.55 μV were classified as "negative responders" ($N = 34$), those who had a positive-going mean amplitude difference that exceeded 1.05 μV were classified as "positive responders" ($N = 34$), and those who fell between these two parameters were classified as "intermediate responders" ($N = 35$). In the next set of analyses, we run separate condition X region analyses for each group. The groups did not differ in their language production and comprehension at 9 months, age at testing, gender, SES, or the number of accepted trials across conditions (see Appendix S2). They also did not significantly differ in their vocabulary at 15 months, reflecting the fact that variation in the segmentation response did not perfectly correlate with vocabulary at 15 months. However, they do differ in the frequencies with which they fall into the low- and high-vocabulary classes. Specifically, negative responders are more likely to be in the high-vocabulary group (65%), intermediate responders are evenly split between the two groups (55% in high-vocabulary group), and positive responders are more likely to be in the low-vocabulary group (70%) ($\chi^2 = 7.7$, $df = 2$, $p = .021$, $\phi = .289$; see Appendix S2). This finding, along with the fact that there have been consistent observations of polarity shifts from positivities to negativities across development (Kooijman et al., 2013; Kudo, Nonaka, Mizuno, Mizuno, & Okanoya, 2011; Männel & Friederici, 2013; Rivera-Gaxiola, Silva-Pereyra, & Kuhl, 2005), means that exploring these differences may deliver new insights into the developmental course of speech segmentation.

One potential objection to dividing the sample into subgroups based on the data we intend to analyze is that we are guaranteed to find effects, at least between the DF and UF conditions. We therefore checked our analyses against new groupings, by taking advantage of the fact that we had twice as many trials in the DF condition. We created ERPs corresponding to every odd DF trial and to every even DF trial, respectively. Next, we created three subgroups based on the difference between the *odd* DF trial and the UF trials. We then tested the subgroups on condition

differences, in which we now relied on the *even* DF trials as our dependent measure. Our main results largely replicated, and there was good split-half reliability in the DF–UF difference score (Spearman-Brown coefficient = .71, Guttman Split-Half = .69, see Appendix S3). We present here the analyses that contain all DF trials.

Negative responders

Event-related potential waveforms for each condition across each quadrant for the negative responders subgroup are shown in Figure 3. A 3 (condition) × 4 (quadrant) way repeated-measures ANOVA revealed a significant main effect of condition [$F(1.62, 85.55) = 14.55$, $p < .001$, $\eta_p^2 = .306$], which was subsumed by a significant condition x quadrant interaction ($F(3.38, 111.62) = 3.31$, $p = .018$, $\eta_p^2 = .091$). Separate follow-up ANOVAs were conducted on each region to identify the source of the interaction. Table 2 lists the results.

From Table 2, we see that the strongest effect of condition occurs over the left frontal region, with both the DF and IF conditions being significantly more negative relative to the UF condition, suggesting that the children in the negative responders group distinguished between all three conditions. There were significant main effects of condition in the right frontal and parietal regions. In both cases, post hoc analyses showed significant differences between the DF and UF conditions.

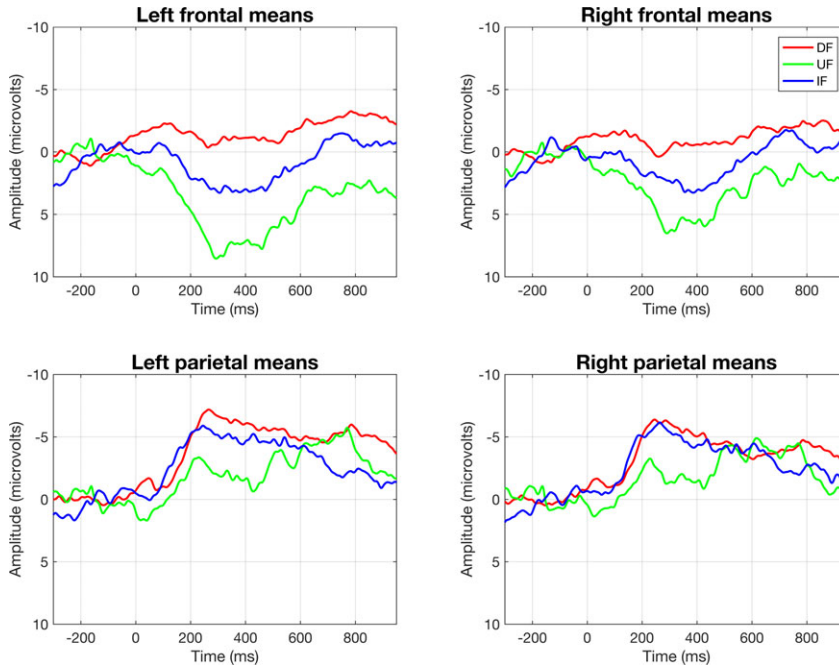

Figure 3 Grand average event-related potential for negative responder group ($n = 34$). Direct familiar (red line), indirect familiar (blue line), and unfamiliar (green line) conditions at the Left Frontal (LF), Right Frontal (RF), Left Parietal (LP), and Right Parietal (RP) regions. Negativity is plotted upwards.

TABLE 2
ANOVA and Post Hoc Results by Quadrant for Negative Responder Group in 250–500 msec Time Window

Quadrant	*Condition main effect*	*Post hoc comparisons**
Left Frontal	$F(1.31, 43.27) = 28.7$, $p < .001$, $\eta_p^2 = .47$	DF < IF < UF
Right Frontal	$F(2, 66) = 12.61$, $p < .001$, $\eta_p^2 = .28$	DF < IF = UF
Left Parietal	$F(2, 66) = 2.73$, $p = .073$, $\eta_p^2 = .076$	NA
Right Parietal	$F(2, 66) = 4.06$, $p = .022$, $\eta_p^2 = .11$	DF < UF

DF = direct familiar; IF = indirect familiar; UF = unfamiliar.
$N = 34$.
*Post hoc Bonferroni comparisons.

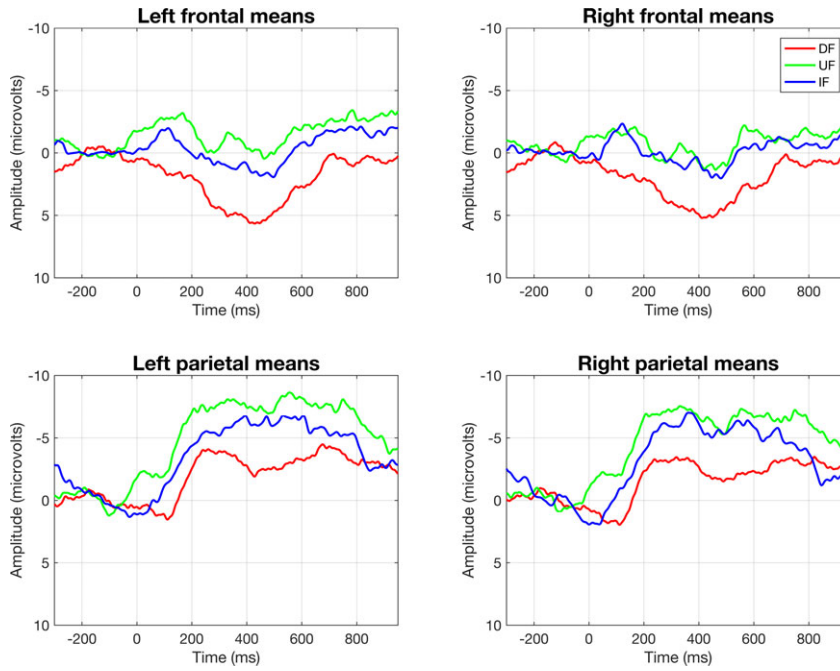

Figure 4 Grand average event-related potential for positive responder group ($n = 34$). Direct familiar (red line), indirect familiar (blue line), and unfamiliar (green line) conditions at the Left Frontal (LF), Right Frontal (RF), Left Parietal (LP), and Right Parietal (RP) regions. Negativity is plotted upwards.

Positive responders

Event-related potential waveforms for each condition across each quadrant for the Positive responders subgroup are shown in Figure 4. A 3 (condition) × 4 (quadrant) way repeated-measures ANOVA revealed a significant main effect of condition ($F(2, 66) = 12.66$, $p < .001$, $\eta_p^2 = .277$). The condition x quadrant interaction was not significant [$F(3.41, 112.43) = .379$, $p = .79$, $\eta_p^2 = .011$], indicating an effect of condition across all four quadrants. Post hoc Bonferroni comparisons across conditions but averaged across quadrants

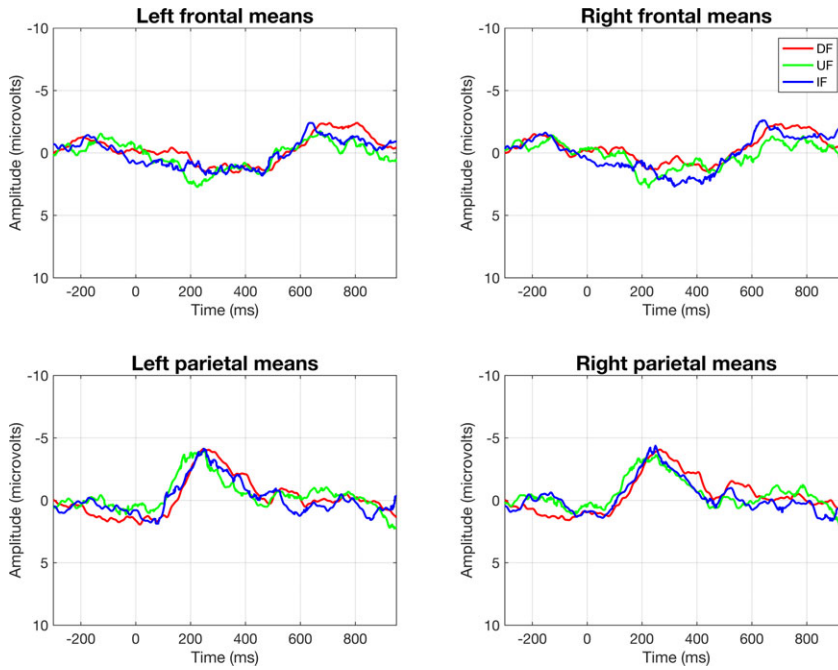

Figure 5 Grand average event-related potential for intermediate responder group (n = 35). Direct familiar (red line), indirect familiar (blue line), and unfamiliar (green line) conditions at the Left Frontal (LF), Right Frontal (RF), Left Parietal (LP), and Right Parietal (RP) regions. Negativity is plotted upwards. Shaded areas denote ± 1*SE*.

revealed the DF condition to be significantly more positive than both the IF (p = .008) and UF conditions (p < .001), while the IF and UF conditions did not differ (p = .65).

Intermediate responders

Event-related potential waveforms for each condition across each quadrant for the Intermediate responders subgroup are shown in Figure 5. A 3 (condition) × 4 (quadrant) way repeated-measures ANOVA revealed no significant effects. [Main effect of condition: $F(1.65, 56.13) = 1.22$, $p = .301$, $\eta_p^2 = .035$; Condition x Quadrant interaction: $F(3.83, 130.36) = .71$, $p = .58$, $\eta_p^2 = .020$].

Event-related potential intra-individual variability and group membership

Whereas the relative negativity and positivity have been previously described (Junge, Kooijman, et al., 2012; Kooijman et al., 2013), no prior study has identified a group of intermediate responders whose ERP record shows little differentiation across conditions. This no doubt reflects the smaller sample sizes used in previous studies, where subgroups were created based on the sign of the children's difference score around zero (i.e., positive or negative, Junge, Kooijman, et al., 2012). If children in the intermediate subgroup actually produce *both* positive and negative responses, varying across individual trials, then we should expect to see greater variability in their ERP signal

compared to the negative and positive responder groups. We tested for this by analyzing each participant's mean standard deviation of their ERP waveform in a 3 (condition) × 4 (quadrant) × 3 (group) way repeated-measures ANOVA. The main effect of group was not significant [$F(2, 99) = .57$, $p = .57$, $\eta_p^2 = .011$], and nor were there any significant interactions involving the variable of group (all Fs < 1.55). These results suggest that the intermediate response represents an individually consistent response category in this 250–500 msec time window.

Another possibility is that the intermediate subgroup shows a recognition response outside the 250–500 msec time window (e.g., in the 600–800 msec window where the overall ERP record showed a second peak). A 3 (condition) × 4 (region) ANOVA on the 600–800 msec time window revealed no significant effects for condition [$F(1.7, 57.86) = 1.58$, $p = .213$, $\eta_p^2 = .044$] or for the condition x region interaction [$F(3.94, 133.93) = 1.02$, $p = .414$, $\eta_p^2 = .029$].

However, there was one property of the intermediate responders subgroup that suggests heterogeneity in responding across the two time windows. As the overall ERP data show a late negativity in the 600–800 msec time region, we allocated the children to responder groups in this time region in the same way as we did for the 250–500 msec time window (i.e., via a ternary split across DF–UF). Table 3 shows group membership across the two time windows.

It can be seen that group membership is fairly stable for the negative and positive responder subgroups. In the case of negative responders, a child who shows a relative negativity between the DF and UF conditions in the 250–500 msec time window has a chance of approximately 65% to also show the same ERP response during the 600–800 msec time window. For positive responders, same-group membership across windows is at 59%. Importantly, being in either of these subgroups in the first time window means there is very little chance of a child producing the opposite response in the second time window: Only 3 of 34 children move from negativity to positivity, and 1 of 34 from positivity to negativity. In contrast, membership in the intermediate responders subgroup in the first time window does not predict subgroup membership in the second time window; rather it is associated with a wider distribution of responses. This difference between the negative and positive responder subgroups, on the one hand, and the intermediate responder subgroup, on the other, is underscored by a significant association for subgroup membership across time windows ($\chi^2 = 33.45$, $df = 4$, $p < .001$, $\phi = .57$).

Therefore, the intermediate responders subgroup contains a subset of participants who show the mature relative negativity in the later 600–800 msec time window; a

TABLE 3
Cross-tabulation of ERP Subgroup Membership in 250–500 and 600–800 msec Time Windows

	ERP responder subgroup (600–800 msec)			
ERP responder subgroup (250–500 msec)	*Negative*	*Intermediate*	*Positive*	*Total*
Negative	22	9	3	34
Intermediate	11	13	11	35
Positive	1	13	20	34
Total	34	35	34	103

ERP = event-related potential.

subset who show the immature relative positivity; and a subset who continue to show no differences. Repeated-measures 3 (condition) × 4 (region) ANOVAs within each subgroup confirm these assignments. For the subset with a late relative negativity ($N = 11$), there was a significant main effect of condition [$F(1.18, 11.82) = 5.59$, $p = .032$, $\eta_p^2 = .359$], and post hoc Bonferroni tests showed the DF condition (M μV = −2.65) to be significantly more negative than the UF condition (M μV = 3.59), $p < .001$, with no other significant effects. The opposite pattern was observed for the subset with a late positivity ($N = 11$): a marginally significant main effect of condition [$F(2, 20) = 3.45$, $p = .052$, $\eta_p^2 = .256$], with mean DF amplitude (M μV = −.597) significantly more positive than mean UF amplitude (M μV = −3.73), $p = .026$), and no other effects significant. For the subset with a continued intermediate response pattern in the later window, no significant effects appeared at all.

DISCUSSION

Our study has addressed the nature of individual differences in infants' speech segmentation skills as measured using ERPs, and the predictive value of such differences for later language skills. Following past research that identified ERP signatures of both immature and mature segmentation in infants learning Dutch (Junge, Kooijman, et al., 2012; Kooijman et al., 2013), we predicted and found significant individual differences in our 9-month-old English acquirers. ERPs were longitudinally associated with the children's vocabulary development even after controlling for several potentially significant predictors of vocabulary growth, showing that the previously attested simple correlations between segmentation and vocabulary represent a developmentally significant effect. Because of the size of our participant group—several times larger than prior studies indicating that language skills might be predicted from segmentation ERPs—we were able to document wide variation in the developmental stage that English-acquiring infants have achieved by 9 months. Under the assumption that past categorizations of immature ("positive") and mature ("negative") ERP signatures of speech segmentation reflect points on a continuum of responding and therefore of abilities, we predicted and found an intermediate response group, whose patterns of responding in the late 600–800 msec time window suggested they may be moving from a clear positive to a clear negative signature. This variation in responding provides a unique view of the developmental course involved in achieving the ability to segment a word form from running speech, and hence to begin the task of building a store of words.

Our negative responder group not only distinguished between the DF and UF stimuli, but they also distinguished IF from UF, effectively recognizing forms that had been presented 10–12 experimental items earlier. Note that such repetition effects are, unsurprisingly, observed in adult native speakers' ERPs (Snijders et al., 2007). Unlike adult native speakers, however, our 9-month-olds would not have lexical entries for words such as *jester* or *sparrow*. Therefore, the data from the negative responders suggests that some 9-month-olds are already skilled at tracking words in running speech and can recognize previously mentioned forms even over delays of approximately 1 min—because not only did these infants segment these words from the speech signal, they also constructed memory traces of them that were stable enough a minute later to elicit a negativity (in comparison with the response to a completely new word, i.e., UF).

Creating a memory trace for a word form is the first step toward building a vocabulary. It is clear from ERP studies of both segmentation and lexical access that a relative negativity is in place where a mature response is observed (e.g., Borgström, von Koss Torkildsen, & Lindgren, 2015; Friedrich & Friederici, 2011; Goyet et al., 2010; Junge, Cutler, & Hagoort, 2012; Junge, Kooijman, et al., 2012; Kooijman et al., 2005, 2013; Männel & Friederici, 2013; Mills et al., 1993, 1997; Thierry et al., 2003). Negative-going responses suggest a listener in possession of language-specific knowledge sufficient to identify candidate words in speech.

In contrast to the negative responder group, the positive and intermediate responder groups' segmentation skills appear less advanced (and their subsequent vocabulary skills were typically lower). While there is evidence that most children in the positive and intermediate groups do distinguish between immediately familiar (DF) and totally unfamiliar (UF) words, their response is qualitatively different (i.e., a relative positivity) or their recognition is delayed (i.e., a relative negativity in the 600–800 msec time window), or both.

Positive-going responses at earlier stages of development have also been observed in prior studies (e.g., Kooijman et al., 2013; Kudo et al., 2011; Männel & Friederici, 2013; Rivera-Gaxiola, Silva-Pereyra et al., 2005), and early infant ERP signals have frequently been reported to be positive-going (e.g., Dehaene-Lambertz & Baillet, 1998; Dehaene-Lambertz & Dehaene, 1994; Leppäanen, Pihko, Eklund, & Lyytinen, 1999; Teinonen, Fellman, Näätänen, Alku, & Huotilainen, 2009). Rivera-Gaxiola, Silva-Pereyra et al. linked a relative positivity to a greater reliance on acoustic properties of the signal at an earlier stage of development, in contrast to the availability of language-specific phonological representations at later stages, with a change in polarity representing the transition from early stage to late.

In our sample, we see this transition in action. Whereas children who were clearly a negative or a positive responder in the 250–500 msec time window tended to maintain their response type or show no effect in the later 600–800 msec time window, the intermediate group did not. In the earlier window, their responses were neither clearly positive nor negative, but in the later window, two-thirds of them had a response in either the positive or negative range. In this case, their response to the speech input is somewhat delayed, but can end up being more of an immature or more of a mature response. Recall that our variability analysis showed that this group was not simply combining some clear positive with some clear negative responses. Rather, we view this group as in a stage in which they are heading toward, though have not yet attained, response maturity.

We note here that, in our data, the change in polarity is due to changes in infants' responses to *both* familiar (i.e., DF) and unfamiliar (i.e., UF) words. The positive responder group's response to unfamiliar words was a negative-going waveform, whereas for the negative responder group, it was positive-going. The groups differed in the opposite direction when processing familiar words. This same trend can be observed in infants acquiring Dutch (Kooijman et al., 2013). Männel and Friederici (2013) also observed a positive- to negative-going response to unfamiliar words in German-acquiring infants aged 6–12 months. Thus, the UF condition effectively constitutes a moving baseline, or, to put it in other words: It is not the single response (e.g., to recognized tokens, or to any token) that alters, it is the entire distribution of responses. There is a shift in the children's processing response as a whole, which we argue is a different processing mode. In our population, about a third of the children

had succeeded in acquiring this processing mode, about a third had not, and the other third was in a position, we suggest, to acquire it very soon.

This account is compatible with the suggestions of Männel and Friederici (2013) and of Rivera-Gaxiola, Klarman, et al. (2005). We suggest that it would be appropriate, however, to take further tentative steps toward an explanatory framework. What might be a processing change so profound that it would induce the young brain to enter a wholly different processing mode? The answer is rather obvious: It is a lexicon. For recognition of spoken language to take off and become part of a communication system, it is necessary that word forms not only be extracted from speech and stored for future recognition, but also that they are paired with meaning. Then, recognition of a familiar form triggers not only a reaction to the acoustic properties of a spoken input, it also releases an associated representation of the word's meaning. This change in the complexity of processing surely warrants the description of a shift to a more advanced mode, and the relative left frontal negativity appears to be its signature. That is, we suggest that the emergence of the relative negativity marks a significant transition to a developmental state where children can not only extract the formal properties words from running speech, but crucially understand that words have a referential function. Note that we are not claiming here that our negative responders inferred meaning during the experiment, but instead that their group membership means that this is within their capabilities. That is, negative responders are developing a mature, online *lexical* mode of processing.

Evidence in support of this interpretation comes from the left-lateralized nature of the IF-UF difference found in the negative responders subgroup. Unlike the DF–UF difference, which was broadly distributed (in both the negative responders and the positive responders), the IF–UF difference was only observed in the left frontal region. This may reflect the gradual left lateralization of language. Recall also that this is the same region that other studies, with smaller samples, use to differentiate positive responders from negative responders. Although the developmental evidence base is sparse, neuroimaging data suggest that lateralization is a gradual and developmental process (Minagawa-Kawai, Cristià, & Dupoux, 2011). That is, what appears to be a general signal-driven preference in the left hemisphere to process acoustic information gradually becomes language-specific through exposure to the input. Importantly, this development seems to be subject to individual differences. For instance, in a study investigating the lateralization of language in 4- to 11-month-olds using near-infrared spectroscopy, Fava, Hull, Baumbauer, and Bortfield (2014) identified a group of children that had a rapid and left-lateralized response to speech in comparison with musical stimuli, and a second group who showed a slower and less lateralized response to speech. In our study, the distinction between IF and UF conditions requires the recently segmented word form (IF) to be distinguished from an as yet unencountered form (UF), which necessitates the formation of a memory phonetic trace for the IF word that lasts at least several minutes. It is the establishment of these memory traces, which likely depend on the brain regions that will become specialized for language, that mark more mature linguistic processing and hence the emergence of the lexicon.

Of course, the infant is intensely busy with a broad range of cognitive developments across the first year of life (Tomasello, Carpenter, Call, Behne, & Moll, 2005). Within the domain of language acquisition, a long tradition of phoneme discrimination experiments has shown how in the early months discrimination is remarkably good for both the phoneme contrasts that are used in the environmental language, as well as for

those that have not actually been heard in speech input, while by the later months of the first year, there is a transition to an adult-like pattern of better discrimination for native-language (heard) than for foreign (unheard) contrasts (Werker & Lalonde, 1988; Werker & Tees, 1984). The ability to process abstract structure is in place very early on (Ferguson & Waxman, 2015), including for spoken input (Gervain, Macagno, Cogoi, Peña, & Mehler, 2008), to the extent that abstract phonological properties of the environmental language can be perceived and stored long before their role in distinguishing word forms is exploited (Choi, Broersma, & Cutler, 2017).

All of this development culminates in the arrival of the first words in an infant's language use. During the second half of the first year of life, a productive co-operation arises between the processes of learning to match word forms with meaning and learning the exact set of sounds needed for identifying words in the environmental language; learning the set of sounds helps in deciding between new words and alternative pronunciations of known words (Werker & Curtin, 2005), while learning new words and their meanings helps to decide upon what sounds indeed distinguish one word from another (Swingley, 2009; Vihman, 2017). Working together, these processes then underpin the tremendously rapid growth that is characteristic of a vocabulary once it is in operation (Johnson, 2016).

Segmentation of natural word forms from speech is seen as early as 6 months of age (Johnson, Seidl, & Tyler, 2014; Johnson & Tyler, 2010), and meaningful associations may be known by then too (Bergelson & Swingley, 2012; Tincoff & Jusczyk, 1999); precursor responses of object-to-sound mapping in the brain appear even at 3 months (Friedrich & Friederici, 2017). In 7- to 10-month-olds (but not, in general, in 5- to 6-month-olds), the process of segmenting a familiarized word from speech has been shown to recruit the brain structures used in adult phonological short-term memory (Minagawa, Hakuno, Kobayashi, Naoi, & Kojima, 2017). But as our findings show, not all children are populating their vocabulary at the same rate.

Two questions that naturally arise from the results concern (1) the psychological construct underlying the transition from a positive to a negative ERP response, and (2) the factors that may underlie the transition. We are only in a position to speculate what these might be, as work in this area is sparse and the contributing factors to individual differences in language acquisition are many and varied (see Kidd et al., 2018). What is self-evident is that the move to a relative negativity implicates a capacity to create and form memory traces for words in a manner that is explicitly linguistic. Thus, one likely explanation implicates basic memory processes like encoding, storage, and retrieval, mediated by the infants' existing language-specific knowledge (language-specific phonology, but potentially also knowledge derived from building their early vocabulary, Junge, Cutler, et al., 2012).

Our tentative suggestion is therefore that the relative negativity is a memory-based effect, which may be supported by attentional processes. The suggestion has some support from research investigating how brain oscillations relate to language and attention. The brain oscillates at different frequencies, which appear to subserve a range of temporally dependent functions. Brito, Fifer, Myers, Elliot, and Noble (2016) reported that variation in resting EEG power within the gamma range (24–48 Hz) at birth was correlated with language comprehension at 15 months. Interestingly, gamma oscillations are thought to reflect the active maintenance of object representations in memory in infants (Kaufman, Csibra, & Johnson, 2005) and have been linked in adults to both attention (Müller, Gruber, & Keil, 2000; Ray, Niebur, Hsaio, Sinai, & Crone, 2008) and language

(e.g., Pulvermüller et al., 1996). Specifically, the gamma range is likely to be implicated in the important early building blocks of linguistic representations—phonemic perception (Meyer, 2017). Clearly more than an ability to discriminate between phonemes is required to successfully segment continuous speech, which requires the infant to identify syllables and the relationships between them. Thus, it is likely that other processes also contribute to individual variability in segmentation, such as differences in sensitivity to syllabic stress and statistical learning (Saffran & Kirkham, 2018).

Pinpointing the factors that may lead to the transition is no small task, especially given that it is likely that development is governed by multiple component processes, all of which may will be influenced by genetic (e.g., differences in timing of brain development) and environmental effects (e.g., differences in quantity and quality of input, parent–child interaction). Meanwhile, our discovery of substantial individual differences in the development of word form segmentation and recognition has consequences for research even in the near future. As described in the introduction, longitudinal studies of individual differences in language acquisition are all too rare; and the present study was set up as part of a longer-term initiative. We will thus be able to establish whether the patterns that we have found to be associated with vocabulary knowledge at 15 months continue to predict language skills as the children in our sample grow older. To date, prospective studies of early and late performance have principally involved vocabulary size, which is both relatively easy to assess and solidly established as a measure on which considerable individual variability displays itself. Comprehension or production measures of vocabulary knowledge have been associated with earlier vowel discrimination scores (Tsao et al., 2004), with segmentation prowess (Newman et al., 2006; Singh et al., 2012), and with many ERP measures (Junge & Cutler, 2014; Junge, Kooijman, et al. 2012; Kooijman et al., 2013; Molfese & Molfese, 1985, 1997; Rivera-Gaxiola, Klarman, et al., 2005). Not everything correlates with later performance (for instance, language discrimination scores do not; Newman et al., 2006), and not every predictive factor remains predictive (e.g., segmentation ERPs of the kind examined here do not predict vocabulary knowledge at age 5 years; Junge & Cutler, 2014). Thus, there is still a lot to be learned about the predictive trajectory (and, as suggested in the Introduction, that could also include earlier measures than the ones taken here, including aspects of the speech input to which infants are exposed, cf. Newman et al., 2016).

Our data offer a finer-grain view of the development of lexical processing skills, in particular in that they appear to differ in what may be the most important reorganization of language processing in that developmental period. Future work with this same large participant group will tell whether they can also offer a finer-grain view of how lexical processing interfaces with other language skills, as well as for how long the earliness of that reorganization continues to show its effects.

ACKNOWLEDGMENTS

We would like to thank all the infants and families who participated, and Deborah Apthorp for help with creating the ERP plots. The research was supported by the ARC Centre of Excellence for the Dynamics of Language (CE40100041, CIs: Cutler & Kidd), and a Major Equipment Grant awarded to Kidd by The Australian National University. We declare no conflict of interests with regard to the funding source for this study.

REFERENCES

Aslin, R. N., Woodward, J. Z., LaMendola, N. P., & Bever, T. G. (1996). Models of word segmentation in fluent maternal speech to infants. In J. Morgan, & K. Demuth (Eds.), *Signal to syntax: Bootstrapping from speech to grammar in early acquisition* (pp. 117–134). Mahwah, NJ: Lawrence Erlbaum.

Bates, E., Dale, P., & Thal, D. (1995). Individual differences and their implications for theories of language development. In P. Fletcher, & B. MacWhinney (Eds.), *Handbook of child language* (pp. 96–151). Oxford, UK: Basil Blackwell.

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

Borgström, K., von Koss Torkildsen, J., & Lindgren, M. (2015). Substantial gains in word learning ability between 20 and 24 months: A longitudinal ERP study. *Brain and Language*, *149*, 33–45.

Brito, N. H., Fifer, W. P., Myers, M. M., Elliot, A. J., & Noble, K. G. (2016). Associations among family socioeconomic status, EEG power at birth, and cognitive skills during infancy. *Developmental Cognitive Neuroscience*, *19*, 144–151.

Choi, J., Broersma, M., & Cutler, A. (2017). Early phonology revealed by international adoptees' birth language retention. *Proceedings of the National Academy of Sciences of the United States of America*, *114*, 7307–7312.

Cristia, A., Seidl, A., Junge, C., Soderstrom, M., & Hagoort, P. (2014). Predicting individual variation in language from infant speech perception measures. *Child Development*, *85*, 1330–1345.

Cristia, A., Seidl, A., Singh, L., & Houston, D. (2016). Test-retest reliability in infant speech perception tasks. *Infancy*, *21*, 648–667.

Dehaene-Lambertz, G., & Baillet, S. (1998). A phonological representation in the infant brain. *NeuroReport*, *9*, 1885–1888.

Dehaene-Lambertz, G., & Dehaene, S. (1994). Speed and cerebral correlates of syllable discrimination in infants. *Nature*, *370*, 292–295.

Fava, E., Hull, R., Baumbauer, K., & Bortfield, H. (2014). Hemodynamic responses to speech and music in preverbal infants. *Child Neuropsychology*, *20*, 430–448.

Fenson, L., Bates, E., Dale, P. S., Marchman, V. A., Reznick, J. S., & Thal, D. J. (2007). *MacArthur-Bates communicative development inventories* (2nd ed). Baltimire, MD: Brookes Publishing.

Fenson, L., Dale, P. S., Reznick, J. S., Bates, E., Thal, D. J., & Pethick, S. J. (1994). Variability in early communicative development. *Monographs for the Society for Research in Child Development*, *59*, 1–185.

Ferguson, B., & Waxman, S. R. (2015). Visual abstract rule learning by 3- and 4-month-old infants. In D. C. Noelle, R. Dale, A. S. Warlaumont, J. Yoshimi, T. Matlock, C. D. Jennings, & P. P. Maglio (Eds.), *Proceedings of the 37th annual meeting of the cognitive science society* (pp. 692–697). Austin, TX: Cognitive Science Society.

Frank, M., Bergelson, E., Bergmann, C., Cristia, A., Floccia, C., Gervain, J., . . . Yurovsky, D. (2017). A collaborative approach to infant research: Promoting reproducibility, best practices, and theory-building. *Infancy*, *22*, 421–435.

Frank, M. C., Braginsky, M., Yurovsky, D., & Marchman, V. A. (2017). Wordbank: An open repository for developmental vocabulary data. *Journal of Child Language*, *44*, 677–694.

Friedrich, M., & Friederici, A. D. (2011). Word learning in 6-month-olds: Fast encoding-weak retention. *Journal of Cognitive Neuroscience*, *23*, 3228–3240.

Friedrich, M., & Friederici, A. D. (2017). The origins of word learning: Brain responses of 3-month-olds indicate their rapid association of objects and words. *Developmental Science*, *20*, e12357.

Gervain, J., Macagno, F., Cogoi, S., Peña, M., & Mehler, J. (2008). The neonate brain detects speech structure. *Proceedings of the National Academy of Sciences of the United States of America*, *105*, 14222–14227.

Goyet, L., de Schonen, S., & Nazzi, T. (2010). Words and syllables in fluent speech segmentation by French-learning infants: An ERP study. *Brain Research*, *1332*, 75–89.

Gross, J., Baillet, S., Barnes, G. R., Henson, R. N., Hillebrand, A., Jensen, O., & Schoffelen, J. M. (2013). Good practice for conducting and reporting MEG research. *NeuroImage*, *65*, 349–363.

Hoff, E., Quinn, J. M., & Giguere, D. (2017). What explains the correlation between growth in vocabulary and grammar? New evidence from latent change score analyses of simultaneous bilingual development. *Developmental Science*, *21*, e12536.

Johnson, E. K. (2016). Constructing a proto-lexicon: An integrative view of infant language development. *Annual Review of Linguistics*, *2*, 391–412.

Johnson, E. K., Seidl, A., & Tyler, M. D. (2014). The edge factor in early word segmentation: Utterance-level prosody enables word form extraction by 6-month-olds. *PLoS One*, *9*, e83546.

Johnson, E. K., & Tyler, M. D. (2010). Testing the limits of statistical learning for word segmentation. *Developmental Science*, *13*, 339–345.

Junge, C., & Cutler, A. (2014). Early word recognition and later language skills. *Brain Sciences*, *4*, 532–559.

Junge, C., Cutler, A., & Hagoort, P. (2012). Electrophysiological evidence of early word learning. *Neuropsychologia*, *50*, 3702–3712.

Junge, C., Cutler, A., & Hagoort, P. (2014). Successful word recognition by 10-month-olds given continuous speech both at initial exposure and test. *Infancy*, *19*, 179–193.

Junge, C., Kooijman, V., Hagoort, P., & Cutler, A. (2012). Rapid recognition at 10 months as a predictor of language development. *Developmental Science*, *15*, 463–473.

Jusczyk, P. W., & Aslin, R. N. (1995). Infants' detection of the sound patterns of words in fluent speech. *Cognitive Psychology*, *29*, 1–23.

Kalashnikova, M., Schwarz, I., & Burnham, D. (2016). Australian English communicative development inventory. *First Language*, *36*, 407–427.

Kaufman, J., Csibra, G., & Johnson, M. H. (2005). Oscillatory activity in the infant brain reflects object maintenance. *Proceedings of the National Academy of Sciences of the United States of America*, *102*, 1571–15247.

Kidd, E., Donnelly, S., & Christiansen, M. H. (2018). Individual differences in langauge acquisition and language processing. *Trends in Cognitive Sciences*, *22*, 152–169.

Kooijman, V. K., Hagoort, P., & Cutler, A. (2005). Electrophysiological evidence for prelinguistic infants' word recognition in continuous speech. *Cognitive Brain Research*, *24*, 109–116.

Kooijman, V. K., Junge, C., Johnson, E. K., Hagoort, P., & Cutler, A. (2013). Predictive brain signals of linguistic development. *Frontiers in Psychology*, *4*, 25.

Kudo, N., Nonaka, Y., Mizuno, N., Mizuno, K., & Okanoya, K. (2011). On-line statistical segmentation of a non-speech auditory stream in neonates as demonstrated by event-related brain potentials. *Developmental Science*, *14*, 1100–1106.

Leppäanen, P. H. Y., Pihko, E., Eklund, K. M., & Lyytinen, H. (1999). Cortical responses of infants with and without a genetic risk for dyslexia: II group effects. *NeuroReport*, *10*, 969–973.

Luck, S. J. (2010). Is it legitimate to compare conditions with different numbers of trials? Retrieved from http://www.erpinfo.org/uploads/5/8/4/6/58469631/mean_peak_noise.pdf

Männel, C., & Friederici, A. D. (2013). Accentuate or repeat? Brain signatures of developmental periods in infant word recognition. *Cortex*, *49*, 2788–2798.

Marchman, V. A., & Bates, E. (1994). Continuity in lexical and morphological development: A test of the critical mass hypothesis. *Journal of Child Language*, *21*, 339–366.

Meyer, L. (2017). The neural oscillations of speech processing and language comprehension: State of the art and emerging mechanisms. *European Journal of Neuroscience*. https://doi.org/10.1111/ejn.13748.

Mills, D. L., Coffey-Corina, S., & Neville, H. J. (1993). Language acquisition and cerebral specialisation in 20-month-old infants. *Journal of Cognitive Neuroscience*, *5*, 317–334.

Mills, D. L., Coffey-Corina, S., & Neville, H. J. (1997). Language comprehension and cerebral specialisation from 13 to 20 months. *Developmental Neuropsychology*, *13*, 397–445.

Mills, D., Conboy, B. T., & Paton, C. (2005). How learning new words shapes the organization of the infant brain. In L. Namy (Ed.), *Symbol use and symbolic representation* (pp. 123–153). Mahwah, NJ: Lawrence Erlbaum.

Minagawa, Y., Hakuno, Y., Kobayashi, A., Naoi, N., & Kojima, S. (2017). Infant word segmentation recruits the cerebral network of phonological short-term memory. *Brain and Language*, *170*, 39–49.

Minagawa-Kawai, Y., Cristià, A., & Dupoux, E. (2011). Cerebral lateralization and early speech acquisition: A developmental scenario. *Developmental Cognitive Neuroscience*, *1*, 217–232.

Molfese, D. L., & Molfese, V. J. (1985). Electrophysiological indices of auditory discrimination in newborn infants: The bases for predicting later language development? *Infant Behavior and Development*, *8*, 197–211.

Molfese, D. L., & Molfese, V. J. (1997). Discrimination of language skills at five years of age using event-related potentials recorded at birth. *Developmental Neuropsychology*, *13*, 135–156.

Müller, M. M., Gruber, T., & Keil, A. (2000). Modulation of induced gamma band activity in the human EEG by attention and visual information processing. *International Journal of Psychophysiology*, *38*, 283–299.

Newman, R., Bernstein Ratner, N., Jusczyk, A., Jusczyk, P., & Dow, K. (2006). Infants' early ability to segment the conversational speech signal predicts later language development: A retrospective analysis. *Developmental Psychology*, *42*, 643–655.

Newman, R., Rowe, M., & Bernstein Ratner, N. (2016). Input and uptake at 7 months predicts toddler vocabulary: The role of child-directed speech and infant processing skills in language development. *Journal of Child Language*, *43*, 1158–1173.

Oakes, L. (2017). Sample size, statistical power, and false conclusions in infant looking-time research. *Infancy*, *22*, 436–469.

Open Science Collaboration (2015). Estimating the reproducibility of psychological science. *Science*, *349* (6251), aac4716

Pulvermüller, F., Eulitz, C., Pantev, C., Mohr, B., Feige, B., Lutzenberger, W., & Birbaumer, N. (1996). High-frequency cortical responses reflect lexical processing: An MEG study. *Electroencephalography and Clinical Neurophysiology*, *98*, 76–85.

Ray, S., Niebur, E., Hsaio, S. S., Sinai, A., & Crone, N. E. (2008). High frequency gamma activity (80–150 Hz) is increased in human cortex during selective attention. *Clinical Neurophysiology*, *119*, 116–133.

Rivera-Gaxiola, M., Klarman, L., Garcia-Sierra, A., & Kuhl, P. K. (2005). Neural patterns to speech and vocabulary growth in American infants. *NeuroReport*, *16*, 495–498.

Rivera-Gaxiola, M., Silva-Pereyra, J., & Kuhl, P. K. (2005). Brain potentials to native and nonnative contrast in 7 and 11 month old American infants. *Developmental Science*, *8*, 162–172.

Saffran, J. R., & Kirkham, N. Z. (2018). Infant statistical learning. *Annual Review of Psychology*, *69*, 181–203.

Saffran, J. R., Werker, J., & Werner, L. (2006). The infant's auditory world: Hearing, speech, and the beginnings of language. In R. Siegler, & D. Kuhn (Eds.), *Handbook of child development* (pp. 58–108). New York, NY: Wiley.

Seidl, A., & Johnson, E. K. (2006). Infant word segmentation revisited: Edge alignment facilitates target extraction. *Developmental Science*, *9*, 566–574.

Singh, L., Reznick, J. S., & Xuehua, L. (2012). Infant word segmentation and childhood vocabulary development: A longitudinal analysis. *Developmental Science*, *15*, 482–495.

Snijders, T., Kooijman, V., Cutler, A., & Hagoort, P. (2007). Neurophysiological evidence of delayed segmentation in a foreign language. *Brain Research*, *1178*, 106–113.

Squires, J., & Bricker, D. (2009). *Ages & Stages Questionnaires. 3rd edition (ASQ-3): A parent-completed child-monitoring system*. Baltimore, MD: Paul Brookes.

Stets, M., Burt, M., & Reid, V. M. (2013). Infants need more variety – Increased data acquisition with reduced participant attrition in infant ERP studies. *Frontiers in Psychology*, *4*, 117.

Swingley, D. (2009). Contributions of infant word learning to language development. *Philosophical Transactions of the Royal Society of London B Biological Sciences*, *364*, 3617–3632.

Teinonen, T., Fellman, V., Näätänen, R., Alku, P., & Huotilainen, M. (2009). Statistical language learning in neonates revealed by event-related brain potentials. *BMC Neuroscience*, *10*(1), 21.

Thierry, G., Vihman, M. M., & Roberts, M. V. (2003). Familiar words capture attention in 11-month-olds in less than 250 ms. *NeuroReport*, *14*, 2307–2310.

Tincoff, R., & Jusczyk, P. W. (1999). Some beginnings of word comprehension in 6-month-olds. *Psychological Science*, *10*, 172–175.

Tomasello, M., Carpenter, M., Call, J., Behne, T., & Moll, H. (2005). Understanding and sharing intentions: The origins of cultural cognition. *Behavioral and Brain Sciences*, *28*, 675–691.

Tsao, F. M., Liu, H. M., & Kuhl, P. K. (2004). Speech perception in infancy predicts language development in the second year of life: A longitudinal study. *Child Development*, *75*, 1067–1084.

van de Weijer, J. (1998). *Language input for word discovery*. PhD Dissertation, Radboud University Nijmegen. MPI Series in Psycholinguistics 9.

Vihman, M. M. (2017). Learning words and learning sounds: Advances in language development. *British Journal of Psychology*, *108*, 1–27.

Vihman, M. M., Thierry, G., Lum, J., Keren-Portnoy, T., & Martin, P. (2007). Onset of word form recognition in English, Welsh and English-Welsh bilingual infants. *Applied Psycholinguistics*, *28*, 475–493.

Werker, J. F., & Curtin, S. (2005). PRIMIR: A developmental framework of infant speech processing. *Language Learning and Development*, *1*, 197–234.

Werker, J. F., & Lalonde, C. E. (1988). Cross-Language speech perception: Initial capabilities and developmental change. *Developmental Psychology*, *24*, 672–683.

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*, 49–63.

SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section at the end of the article.

Appendix S1. Explanation of outliers in MB-CDIs and segmentation ERP difference score.

Appendix S2. Comparisons across ERP groupings on language, age, gender, SES and number of accepted trials in experiment.

Appendix S3. Replication of group-level analyses (with every other trial).