RUNNING HEAD: INHERENT CONSTRAINTS ON ENCODING SPEECH SEQUENCES

On the Edge of Language Acquisition: Inherent Constraints on Encoding Multisyllabic Sequences in the Neonate Brain

Alissa L. Ferry[1]*, Ana Fló[1]*, Perrine Brusini[1], Luigi Cattarossi[2], Francesco Macagno[2], Marina Nespor[1] & Jacques Mehler[1]

[1] Language, Cognition, and Development Laboratory, Scuola Internazional Superiore di Studi Avanzati, Trieste, Italy

[2] Neonatology Unit, Azienda Ospedaliera Santa Maria della Misericordia, Udine, Italy

*co first author

Author Note:

We thank the staff at the Neonatology and Obstetrics Departments of the Azienda Ospedaliera Santa Maria della Misericordia for their assistance and to the parents of the newborn participants for their participation. We would like to thank Silvia Benavides-Varela, Alex Cristia, Susan Hespos, Heather Lucas, and Susana Franck for feedback and discussions on previous versions of the manuscript. The research leading to these results has received funding from the European Research Council under the European Unions' Seventh Framework Programme (FP7/2007-2013)/European research Council Grant Agreement 269502 (PASCAL) (to J.M). Author Contributions: A.L.F, A.F., M.N., and J.M. designed the project; A.L.F., A.F., P.B., F.M., and L.C performed the research; A.F. analyzed the data; A.L.F, A.F., P.B., M.N., and J.M. wrote the paper.

Correspondence concerning this article should be addressed to Alissa Ferry, Cognitive Neuroscience Sector, Scuola Internazionale Superiore di Studi Avanzati, Via Bonomea 265, Trieste, Italy 34136.

Contact: aferry@sissa.it, (+39) 040-378-7613

Research Highlights

- These findings show that newborn infants encode the edges of multisyllabic sequences better than their internal components.
- This research also demonstrates that newborns use subtle prosodic cues in the speech signal to segment multisyllabic sequences, with enhanced encoding of syllables at the edges of subtle prosodic boundaries.
- This study suggests that humans are born with the fundamental mechanisms for encoding the order of syllables in sequences, a skill indispensible for tracking the hierarchical organization of language.

Abstract

To understand language, humans must encode information from rapid, sequential streams of syllables— tracking their order and organizing them into words, phrases, and sentences. We used Near-Infrared Spectroscopy (NIRS) to determine whether human neonates are born with the capacity to track the positions of syllables in multisyllabic sequences. After familiarization with a six-syllable sequence, the neonate brain responded to the change (as shown by an increase in oxy-hemoglobin) when the two edge syllables switched positions but not when two middle syllables switched positions (Experiment 1), indicating that they encoded the syllables at the edges of sequences better than those in the middle. Moreover, when a 25ms pause was inserted between the middle syllables as a segmentation cue, neonates' brains were sensitive to the change (Experiment 2), indicating that subtle cues in speech can signal a boundary, with enhanced encoding of the syllables located at the edges of that boundary. These findings suggest that neonates' brains can encode information from multisyllabic sequences and that this encoding is constrained. Moreover, subtle segmentation cues in a sequence of syllables provide a mechanism to accurately encode positional information from longer sequences. Tracking the order of syllables is necessary to understand language and our results suggest that the foundations for this encoding are present at birth.

Keywords: language acquisition, sequential encoding, NIRS, speech processing, neonates

On the Edge of Language Acquisition: Inherent Constraints on Encoding Multisyllabic Sequences in the Neonate Brain

Introduction

One of the fundamental goals of cognitive science is to understand the mechanisms that underlie infants' rapid acquisition of language and to determine which mechanisms operate from birth and which require environmental experience. Here we investigate whether the ability to encode positional information for items in a sequence, a necessary component of language processing, is one of the mechanisms that constrain language processing from birth. The temporal nature of language requires that sequential information be encoded; in spoken language, syllables are sequentially organized into words, words into phrases, and phrases into sentences. For example, learning that the word *banana* refers to the concept of a banana requires that infants encode the sequence of syllables that make up the word. They must not only learn that the word consists of three syllables /ba/, /na/, and /na/, but also that the three syllables are arranged in a specific order. At the sentence level, the sentence *the dog bites the boy* has an entirely different meaning from the sentence *the boy bites the dog*, while *dog boy the bites the* is completely ungrammatical and meaningless; even though each sentence is constructed using the same words, the position of those words alters the meaning. Here, we investigate whether a signature constraint of general sequential processing in adults— an enhanced encoding of sequence edges— is evident from birth and influences how linguistic sequences are encoded. Specifically, we ask if neonates encode positional information from six-syllabic sequences and if they encode some positions (i.e., syllables

at the edges) more precisely than others (i.e., syllables in the middle), in accordance with the constraints on general sequential processing found in adults.

Since language requires encoding sequential information, sequential processing constraints may influence language processing. Extensive research shows that a fundamental constraint in processing sequential information lies in the fact that edges of a sequence are encoded more precisely than internal positions. Items at the edges of a sequence are better recalled than items in the middle, a phenomenon known as the serial position effect (Ebbinghaus, 2013). This enhanced memory for items at the beginning (primacy effect) and at the end (recency effect) of a sequence is robust, with the same pattern manifesting itself in many domains, including linguistic, visual, and spatial domains (Gupta, Lipinski, Abbs, & Lin, 2005; Hurlstone, Hitch, & Baddeley, 2014). Moreover, when the order of items in a sequence is encoded, only the edge positions appear to be encoded precisely, while all other positions appear to be encoded relative to the edges (Endress, Nespor, & Mehler, 2009; Henson, 1998). In a sequence, such as ABCDE, A is encoded precisely as being in the first position and E in the last position, while BCD are encoded less precisely and only in relation to the edge positions (e.g., B is one place after the first position or three places before the last position (Henson, 1998).

If positional information of sequences is encoded relative to edges, this could explain why linguistic regularities are generally edge-based and edge-based artificial grammars are easily learned (Endress et al., 2009). Across languages, determiners, bound morphemes, and word primary stress are positioned with respect to the edge of a word (Greenberg, 1957; Hayes, 1995; Kager, 1995), suggesting that languages capitalize on this enhanced encoding of edges by placing regularities with respect to the more precisely

encoded edges. This idea is further supported by artificial grammar learning experiments, which demonstrate an enhanced encoding of items located at the edges of sequences (Endress et al., 2009). Adults extract a positional-based regularity (e.g., an immediate repetition of two syllables in a seven-syllable sequence) when that regularity is at the edge of the sequence but not when it is internal (Endress & Mehler, 2009). Adults fail to extract a structural regularity— an AxC pattern with words in which A always predicts C with an irrelevant syllable separating them—from a continuous stream, but succeed when the words are separated by an imperceptible 25ms pause (Peña, Bonatti, Nespor, & Mehler, 2002). This finding shows that when the boundaries were marked by another cue, even if it was very subtle, the regularity at the edge positions was generalized. Combined, the existing research suggests that the reliance on sequential information during speech processing privileges the encoding of the edges, and that cues signaling a boundary within a sequence influence how the sequential information is encoded.

While extensive research has investigated the edge bias in adults (e.g., Hurlstone et al., 2014), less research has investigated how sequential information is encoded early in development and whether there are constraints on sequential processing in infancy that might influence early language acquisition. Existing research on early memory capabilities suggests that newborns are capable of remembering the rhythmic patterns of stories the mother read out loud during the last trimester of pregnancy (DeCasper & Spence, 1986), or short bisyllabic words they heard shortly after birth (Benavides-Varela et al., 2011; Valiante, Barr, Zelazo, Papageorgiou, & Young, 2006). These findings suggest that humans are born with the ability to extract and remember information that is potentially useful for language acquisition. Yet, the extent of infants' ability to extract

information remains an open question. Our question focuses on whether infants encode and remember any information from longer sequences of syllables and whether that encoding is limited by the same constraints on encoding sequences that are evident in older infants and adults.

Early research examining sequential encoding in infants focused primarily on infants' ability to encode sequences of visual stimuli. For example, one study presented 7-month-old infants with a sequence of three faces and found that after a short delay, infants recognized the faces presented first and last, but not the face presented in the middle of the sequence. Other studies have demonstrated long-term memory serial position effects in 3- and 6-month-olds using small visual patterns (Gulya, Sweeney, & Rovee-Collier, 1999; Gulya, Galluccio, Wilk, & Rovee-Collier, 2001; Gulya, Rovee-Collier, Galluccio, & Wilk, 1998). Some research also provides evidence of an enhanced encoding of edges of linguistic stimuli. Eight-month-olds can segment and remember novel nouns from a sentence if the novel noun is the first or last word of the sentence but not if the novel noun is located in the middle of a sentence (Seidl & Johnson, 2006), suggesting an enhanced encoding of the edges of sequences of words constituting sentences. A recent study (Benavides-Varela & Mehler, 2014) demonstrated that 7-month-olds show a more precise encoding of syllables at the edges of a sequence than of syllables in the middle of a sequence; they detected a change when the edge syllables of a five-syllable sequence switched positions but not when two of the internal syllables switched positions, suggesting that the edge syllables were better remembered.

The enhanced encoding of sequence edges appears to emerge before the end of the first year of life, and this is evident with both visual and auditory stimuli. Yet there

are still open questions of whether this edge bias is an inherent constraint on sequential processing and whether this constraint influences how language, a sequentially presented stimulus, is processed from birth. By 7 months infants are already sensitive to some edge-based linguistic regularities (Gervain, Nespor, Mazuka, Horie, & Mehler, 2008; Seidl & Johnson, 2006; Thiessen & Saffran, 2007) and exposure to such robust cues may tune infants to an enhanced encoding of sequence edges. In Experiment 1 we tested whether this edge bias is a processing constraint that is present at birth, prior to extensive exposure to language.

Because the goal of the current research is to investigate whether these constraints influence the encoding of multisyllabic sequences and potentially language acquisition from birth, we focused our experimental design on how neonates encode sequences of syllables. While no studies have examined whether newborns remember information from longer multisyllabic sequences, it has been shown that they can remember bisyllabic sequences after only very brief exposure shortly after birth (Benavides-Varela et al., 2011; Benavides-Varela, Hochmann, Macagno, Nespor, & Mehler, 2012). In Benavides-Varela et al. (2011) researchers used functional Near-Infrared Spectroscopy (NIRS) to measure changes in brain responses during a memory task with neonates. After familiarization with a bisyllabic sequence (e.g., mita), infants who heard a new sequence (e.g., pelu) showed a greater increase in the hemodynamic response than infants who heard the same familiarization sequence, suggesting that newborns' brains can recognize short sequences.

We employed a similar paradigm in which we familiarized neonates with a six-syllable sequence and examined whether the brain responded to a change in the sequence

when either the two edge syllables switched positions (Edge Switch condition) or two internal syllables switched positions (Internal Switch condition). Switching the syllables, rather than replacing them with novel syllables allowed us to determine whether infants were able to encode, not just the syllable's identity, but also information about its position within a sequence.

Experiment 1

To determine if neonates were able to detect the change in the sequence, we employed NIRS (Gervain et al., 2011; Lloyd-Fox, Blasi, & Elwell, 2010), a methodology particularly suitable for neonates. This technique uses near infrared light to measure changes in the concentration of oxygenated hemoglobin (HbO) in the blood vessels in surface cortical regions, allowing for inferences about brain responses to external stimuli. We used a habituation and change detection paradigm. It is a well-documented phenomenon that repetition of the same stimulus leads to a decrease in neural activity and a subsequent change in the stimulus triggers a recovery of neural activity. This pattern is consistently found in neuroimaging studies on newborns and older infants (Dehaene-Lambertz, Hertz-Pannier, et al., 2006; Mahmoudzadeh et al., 2013) and in adults (Dehaene-Lambertz, Dehaene, et al., 2006). In particular, NIRS habituation experiments with infants and neonates often find a relatively broad, bilateral pattern of increased HbO response to a stimulus change after habituation. For example, 3- and 4-month-olds were familiarized to blocks of a syllable (e.g., pa) and showed a broad, bilateral increase (particularly in frontal and temporal regions) in response to the presentation of a block with new syllables (e.g., ba), compared to the presentation of a block of the same syllable (Nakano, Watanabe, Homae, & Taga, 2009). A similar broad, bilateral, fronto-temporal

response was found when neonates were familiarized to blocks of a bisyllabic word and then listened to blocks containing a new word (Benavides-Varela et al., 2011). The current experimental protocol takes advantage of these robust habituation effects, by examining differences in the recovery patterns between test blocks in which the edge syllables of the sequence switched positions and test blocks in which two internal syllables switched positions. We used one silicone pad on each side of the head, with emitters and detectors aligned to form 12 channels on each pad for a total of 24 channels for analysis. (see Fig. 1A). Our block design (Fig. 1B) alternated between six familiarization blocks during which infants heard the repetition of the same sequence twenty times (e.g., simebutal*e*fo), and six test blocks. In test blocks, a modified version of the sequence was repeated five times. For half of the neonates, the two edge syllables changed position (Edge Switch, e.g., **fo**mebutal*e***si**) and for the other half, the third and fourth syllables of the word changed position (Internal Switch, e.g., sime**tabu**l*e*fo). By examining increases in the hemodynamic response during the test blocks, we could determine if infants' brains are sensitive to the change from familiarization to test. We hypothesized that if the positional information for syllables was encoded better at the edges, a switch of the edge syllables would be detected more easily than a switch of the internal syllables. This should lead to a greater hemodynamic response in the Edge Switch condition than in the Internal Switch condition.

Experiment 1: Methods

Participants

In this and in the following experiment, all participants were healthy, full-term neonates born to Italian-speaking mothers. The criteria for inclusion were Apgar score ≥

7 in the first minute and ≥ 8 in the fifth minute, diameter of head ≥ 33.0 cm, and no cefalhematomas. In Experiment 1 the Edge Switch condition included 16 participants (6 females; mean age 2.75 days, range 2-4 days; mean gestational age 38.9 weeks, SD 1.1 weeks; mean weight 3.475 Kg, SD 0.336 Kg), and the Internal Switch condition included 16 participants (10 females; mean age 3.1 days, range 2-4 days; mean gestational age 38.9 weeks, SD 1.1 weeks, mean weight 3.382 Kg, SD 0.359 Kg), who provided data without motion artifacts from at least two of the six test blocks. Additional infants were tested but excluded from the final analyses because of motion artifacts during more than 4 of the test blocks (Edge Switch condition n = 11, Internal Switch condition n = 8), failure to complete the experiment (Edge Switch condition n = 3, Internal Switch condition n = 5), a poor signal due to thick hair (Edge Switch condition n = 6, Internal Switch condition n = 2), or computer error (Edge Switch condition n = 1). This attrition rate is consistent with other studies that use NIRS with neonates (Benavides-Varela et al., 2011; Gervain et al., 2011; Lloyd-Fox et al., 2010). All newborns were recruited from the nursery at the Hospital, Azienda Ospedaliera Santa Maria della Misericordia, in Udine, Italy. Parents provided informed consent. The Ethical Committee of the Scuola Internazinale Superiore di Studi Avanzati approved the study.

Stimuli

All sequences consisted of six different consonant-vowel (CV) syllables. A total of 12 pairs of sequences were used in each condition (See Table 1). The sequences were synthetized using the *it4* Italian female voice of the MBROLA diphone database (Dutoit, Pagel, Pierret, Bataille, & Van der Vrecken, 1996), with phoneme duration of 150 ms and a constant pitch of 200Hz. Sequences were continuous with no pauses between syllables.

All test sequences were created by switching the position of two syllables from the familiarization sequences. In the Edge Switch condition, the positions of the first and last syllables were switched. In the Internal Switch condition the positions of the third and fourth syllables were switched. Each sequence used during the familiarization for one infant was used for the test phase of another infant (e.g., one infant in the Edge Switch condition heard "simebutal*e*fo" during familiarization blocks and "fomebutal*e*si" during the test blocks while another infant heard "fomebutal*e*si" during familiarization blocks and "simebutal*e*fo" during the test blocks). Infants were randomly assigned to a condition and to the pair of sequences that were used during the familiarization and test blocks. Example stimuli are available as Supplementary Material.

Data Recording

Data were recorded using an ETG-4000 NIRS machine (Hitachi Medical Corporation, Tokyo, Japan) that uses two continuous light source wavelengths (695 and 830 nm). The separation between emitters and detectors was 3 cm, the sampling rate was 10 Hz, and total laser power output per fiber was 0.75 mW. Each probe consisted of nine fibers from which five were emitters and four were detectors.

Protocol

The experiment consisted of six sets of familiarization and test blocks. During each familiarization block the same sequence was repeated 20 times, separated by a silence of 0.5-1.5 s. During each test block the test sequence was repeated 5 times, separated by a silence of 0.5-1.5 s. The total duration was 56 s for each familiarization block, and 14 s for each test block. All blocks were separated by periods of silence of random lengths between 25 and 30 s to allow the hemodynamic response to return to baseline. The total

duration of the experiment was 13 minutes.

The neonates were tested while lying in their cribs, asleep or in a state of quiet rest (typical protocol for neonate NIRS experiments (Benavides-Varela et al., 2011; Gervain et al., 2011; Lloyd-Fox et al., 2010)), in a dimly lit sound-attenuated booth. Sound stimuli were presented at approximately 60 dB via two loudspeakers placed on both sides at the feet of the infant's crib at a 30° angle. The speakers were connected to a Macintosh power PC G5 computer that simultaneously operated the NIRS machine and presented the auditory stimuli using PsyScope X software (Cohen, MacWhinney, Flatt, & Provost, 1993). Both the NIRS machine and the computer were placed outside the experimental booth. Two silicon probes (each 7 cm x 9 cm), containing 12 recording points each, were used to keep the optical fibers in place. One probe was placed over the right side of the head (channels 1-12) and the other over the left (channels 13-24), using skull landmarks (the bottom detector was placed above the ear and the probe was kept aligned along the anterior-posterior direction) (Fig. 1B). The positioning was chosen to maximize the recording from fronto-temporal regions. During the testing session, an experimenter controlled the NIRS machine from outside the room, a second experimenter held the probes in place, a medical doctor, blind to the experimental hypotheses, assisted the neonate, and parents were free to remain in the booth or not. An infrared video camera was used to monitor the infant's behavior.

Data Processing

Data were analyzed using custom functions in MATLAB 2012b (*MATLAB and Statistics Toolbox Release 2012b*, n.d.) according to a general protocol used previously in other NIRS experiments (Benavides-Varela et al., 2011, 2012; Peña et al., 2003).

The preprocessing can be divided in three main steps that were designed to convert the signal to a time course of changes in concentrations of oxy-hemoglobin (HbO) and deoxy-hemoglobin (Hb), to extract a hemodynamic response function (HRF) for each block and channel, and to reject specific channels and blocks that did not provide sufficient data due to noise or motion artifacts.

HbO and Hb time series calculation: First, motion artifacts were detected based on intensity changes for each wavelength and corrected using functions of the Homer2 NIRS package (Huppert, Diamond, Franceschini, & Boas, 2009). To detect the motions artifacts we applied the hmrMotionArtfactsByChannel function on the z-scored data for each wavelength. The function computed samples in a moving time window of 0.5s (tMotion parameter) and marked samples in ± 0.3s (tMask parameter) as motion artifacts if either there was a change in the z-scored signal greater than 0.5 (amp_thresh parameter), or the signal deviated from the mean by more than 4 standard deviations (std_thresh parameter). Motion artifacts were corrected using the spline interpolation algorithm (Cooper et al., 2012; Scholkmann, Spichtig, Muehlemann, & Wolf, 2010), by applying the hmrMotionCorrectSpline function (p_spline = 0.99). After motion artifacts were identified, this algorithm independently fit each artifact using cubic spline interpolation and the fit of the motion artifact was subtracted from the signal.[1] Note that some motion artifacts are unable to be corrected. Uncorrected motion artifacts are identified and excluded in a later step of the analysis. Next, intensity was converted to optical density, and variations of HbO and Hb concentrations were calculated from the changes in the

[1] The p_spline parameter, ranging from 0 to 1, determines the smoothing of the fitting and according to Scholkmann et al. (2010) a value of 0.99 is effective.

optical density on the two wavelengths using the algorithm hmrOD2Conc in the Homer2 NIRS package. Then, the data were band-pass filtered between 0.02 Hz and 0.80 Hz to reduce slow systemic physiological hemodynamic fluctuations such as respiratory signals, and blood pressure changes; and high-frequency instrument noise and the fast cardiac oscillations (~1Hz). Finally, the time series for the entire experiment were lineally detrended.

HRF calculation: In calculating the HRF, we focus only on the HbO signal, which has a better signal-to-noise ratio than the Hb signal (Benavides-Varela et al., 2011; Bortfeld, Wruck, & Boas, 2007; Gervain et al., 2011; Lloyd-Fox et al., 2010). The HRFs for each block were extracted form the time series for HbO by cutting from 5 s preceding the onset till 15 s after the offset (-5s to +29s in the case of test blocks). The first 10 s after the offset accounted for the delay in the BOLD response. The mean value during the first and final 5 s were used to calculate a linear baseline trend that was removed from the signal.

HRF rejection: A response for a block and channel was excluded if during that period 1) the changes in intensity showed saturation (a light absorption of less than 1% of the total light), 2) the changes in intensity provided a low signal-to-noise ratio (a ratio between the standard deviation and the mean greater than 5 in a moving window of 5s, usually due to thick hair), or 3) there were fast changes in HbO. Despite the correction of motion artifacts, the correction method did not always properly reconstruct the signal, and in those cases were excluded from the analysis. In order to automatize the detection of fast changes we applied the motion artifact detection algorithm again, but on the HbO signal for each HRF (tMotion = 0.5, std_thresh = 4, amp_thresh = 0.08). In addition, to

check for motion artifacts that may not have been detected by the algorithm, HRFs showing a mean change in HbO greater than 2 standard deviations from the group mean were visually inspected, and compared to annotations made during the experiment and the recorded videos. These HRFs were manually rejected if motion artifacts that were undetected by the algorithm were identified (e.g., HRFs showed fast signal changes that did not reach the criteria set by the algorithm but were located adjacent to channels where motion artifacts were identified, or HRFs coming from the final block of the recording session that showed extremely big and slow changes over all channels as an artifact introduced by the filtering process). Considering all the tested infants (including infants who were ultimately excluded for not providing at least two motion-free blocks) in the Edge Switch condition 49% of the HRFs were automatically rejected, and 8% manually rejected, whereas in the Internal Switch condition 50% of the HRFs were automatically rejected and 6% manually rejected. Blocks with more than 50% rejected channels were excluded. Infants were only included in the analysis if they contributed at least two valid test blocks. Each of the included infants contributed an average of 3.81 blocks (SD = .91) in the Edge Switch condition, and 3.50 (SD = 1.15) in the Internal Switch condition. Due to the fact that the infants' heads are curved and that infants were tested while lying in their cribs, some channels in the posterior regions were more susceptible to saturation and low signal-to-noise resulting in their exclusion from the data analysis. Across the two experiments presented in this paper, channels 8, 11, 22, 23 and 24 were found to be most susceptible (see Figure 1). On each of these channels fewer than 9 out of 16 infants contributed data for that channel in at least one condition. Because these channels contributed data from fewer infants overall than the other channels the main cluster

analyses were conducted both with and without these channels (to preview, similar results were found across both analyses) and the overall mean change analysis for each condition (collapsing across channels) excluded these channels. For the remaining channels an average of 12.53 (SD = 1.78) infants provided data per channel in the Edge Switch condition and 13.00 (SD = 1.45) in the Internal Switch condition.

Statistical Analysis

All the statistical analyses were performed using the average HRF per infant per channel. We performed two types of data analyses, a cluster-based permutation analysis and a standard mean activation analysis.

Cluster Based Permutation Analysis. This analysis was initially developed for EEG and implemented in the Fieldtrip Matlab toolbox (Oostenveld, Fries, Maris, & Schoffelen, 2010) and has previously been applied to NIRS data (Mahmoudzadeh et al., 2013). It is a non-parametric test that enables comparison of the signal changes across two conditions taking into account spatial proximity between channels and temporal proximity within the time-course. Spatio-temporal clusters are identified in which the conditions are significantly different, controlling for multiple comparisons (one comparison for every time bin and every channel). The method is designed to account for the fact that similar responses are expected between samples that are close in time (a sequence of samples comprising the time course for one channel) or space (nearby channels) and are thus not independent; therefore, it retains sufficient statistical power while adequately controlling for the problem of multiple comparisons. In the cluster-based analysis, two-sample t-tests were conducted between each pair of data points (one comparison per each couple (channel, time)). Then cluster candidates were identified by

grouping all temporally and spatially adjacent pairs with a p-value smaller than a chosen threshold (standard value of .05), which did not affect the false discovery rate. Next, cluster-level statistics were calculated for each cluster candidate by summing the t-value from the two-sample t-test for every data point included in a cluster candidate. This value is larger for larger clusters (more t-values added) and for clusters with larger differences between conditions (larger individual t-values). Finally, a permutation analysis evaluated whether this cluster-level statistic was significantly different from chance. A null distribution was obtained by randomizing the conditions and the proportion of random partitions that produce a cluster-level statistic greater than the observed one was the Monte Carlo p-value for the cluster that determines significance.

In Experiment 1 the Edge Switch condition was compared to the Internal Switch condition. The HRFs for each condition, obtained by the pre-processing described before and lasting 34s (from -5s to 29s from the onset) were used. Given that the HRF is intrinsically slow (~10s), the data were smoothed by down-sampling to 1Hz without a loss of temporal resolution. This resulted in 816 pairs of data points to compare (24 channels x 34 samples). The threshold p-value used to select the pairs of samples to build the clusters was .05. Two pairs of samples were considered temporally adjacent if they were consecutive (time difference of 1s, given that the sampling frequency is 1Hz), and spatially adjacent if they were at a distance < 3cm (the neighbors for channel 1 were, channels 3 and 4; for channel 2, channels 4 and 5; for channel 3, channels 1 and 6; for channel 4, channels 1, 2, 6 and 7; for channel 5, channels 2 and 7; for channels 6, channels 3, 4, 8 and 9; for channel 7, channels 4, 5, 9 and 10; for channel 8, channels 6 and 11; for channel 9, channels 6, 7, 11 and 12; for channel 10, channels 7 and 12; for

channel 11, channels 8 and 9; for channel 12, channels 9 and 10; and in an analogous way for the right probe). 1000 randomizations were made in the permutation analysis to obtain the Monte Carlo p-value.

Mean Activation Analysis.

In order to clarify that our findings result from differences in response to the sequence changes, we also compared the overall change in HbO across all channels – excluding the posterior channels that did not provide sufficient reliable signal– between the familiarization blocks and test blocks across conditions. Differences between the test blocks, but not the familiarization blocks would suggest that the neonate brain processed the edge and internal switches differently. We analyzed the mean change in HbO during the familiarization blocks time window (from 0 to 66 seconds from the onset of each block) and test blocks time window (from 0 to 24 seconds from the onset of each block). Both time windows were calculated from the onset of the stimuli to 10s after the offset. We performed a two-way ANOVA with factors of Condition (Edge Switch/ Internal Switch) and Block Type (Familiarization/ Test).

Experiment 1: Results

Cluster Based Permutation Analysis

We performed a cluster based permutation statistical analysis on the hemodynamic response of the HbO signal in order to examine whether infants responded differently in the Edge Switch and Internal Switch conditions across individual channels and across time.

The analysis revealed a greater hemodynamic response for the Edge Switch condition than for the Internal Switch condition in both hemispheres mostly in tempo-

frontal areas, as is demonstrated by the presence of two clusters. A cluster in the right hemisphere included channels 13, 14, 16, 17, 18, 19 and 21 ($p_{cluster\_1} < 0.0001$), and a cluster in the left hemisphere included channels 1, 3, 4, 5, 6, 7 and 9 ($p_{cluster\_2} = 0.004$) (Fig. 2). We also ran the analysis excluding the channels in posterior areas that were more susceptible to exclusion (8, 11, 22, 23 and 24) with similar results (cluster 1 included channels 13, 14, 16, 17, 18, 19 and 21, $p_{cluster\_1} < 0.0001$; and cluster 2 channels 1, 3, 4, 5, 6, 7 and 9, $p_{cluster\_2} = 0.004$). The results demonstrate that neonates in the Edge Switch condition showed a larger increase in HbO during test blocks than infants in the Internal Switch condition.

Mean Activation Analysis

The ANOVA with factors of Condition (Edge Switch/ Internal Switch) and Block Type (Familiarization/ Test) found a main effect of Condition ($F(1, 15) = 15.52$, $p = .0002$), a main effect of Block Type ($F(1, 15) = 13.27$, $p = .0006$), and a significant interaction between Condition and Block Type ($F(1, 15) = 21.20$, $p = .00002$). Post-hoc Tukey-Kramer comparisons revealed that the mean activity in the test block of the Edge Switch condition differ from the Internal Switch condition ($p = 6x10^{-7}$). The Familiarization blocks do not differ between conditions ($p > .05$). Familiarization and Test blocks are different in the Edge Switch ($p = 10^{-6}$) condition, but not in the Internal Switch condition ($p > .05$).

Experiment 1: Discussion

Experiment 1 sought to uncover whether neonates' brains encode positional information for the items at the edges of a sequence better than for items in the middle. The results revealed that after familiarization with a six-syllable sequence, neonates

showed a larger increase in HbO during test blocks in which the two edge syllables switched position than when the two internal syllables switched positions. The brain response showed evidence of detecting the change in the sequence when the edge syllables, but not the internal syllables switched positions, indicating that the edge syllables were better encoded than the internal syllables. While previous results have demonstrated this phenomenon in adults (Endress & Mehler, 2009; Endress et al., 2009; Henson, 1998; Hurlstone et al., 2014) and in older infants (Benavides-Varela & Mehler, 2014), the current findings demonstrate that the constraints on sequential positional encoding are present from birth, before infants have extensive experience with processing sequential information. This suggests that the edge bias is an inherent constraint in sequential processing that influences how neonates process linguistic information.

Here we show that the differential response between the Edge Switch and Internal Switch conditions emerged bilaterally and across broad regions comprising pre-frontal, frontal, and temporal areas. This extended pattern of activation has been observed in neonate novelty detection experiments (Benavides-Varela et al., 2011, 2012) and researchers have speculated about the roles played by these different regions. Both pre-frontal and frontal regions are involved in novelty detection across a range of tasks (Mahmoudzadeh et al., 2013; Nakano et al., 2009) and sensitivity to speech processing in both temporal regions has been demonstrated in young infants and neonates (Benavides-Varela et al., 2011, 2012; Lloyd-Fox et al., 2010; Mahmoudzadeh et al., 2013). The broad, differential response across multiple channels is common in newborn NIRS experiments, and we note that while we can conclude that there are clear differences

between the conditions, conclusions regarding specific regions of activation must be treated with caution.

These findings demonstrate that newborn's brains respond to a change in the edge of a sequence but not to a change in the middle of a sequence, however there are several possible explanations. The Edge Switch test stimuli swapped the positions of the first and last syllables in the sequence. One possible interpretation is that neonates better encode both items at the beginning of a sequence and items at the end of a sequence, detecting that they have switched positions. An alternative interpretation is that neonates better encode, or even only encode the first (or the last) item in a sequence. While previous research suggests that neonates encode at least some information from more than just one item in a sequence (Gervain, Macagno, Cogoi, Peña, & Mehler, 2008; Gervain, Nespor, et al., 2008; Hevia, Izard, Coubart, Spelke, & Streri, 2014; Teinonen, Fellman, Näätänen, Alku, & Huotilainen, 2009), our current methodology cannot discriminate between these interpretations. Regardless of the underlying processing mechanism, the current results indicate that sequence edges (either only one edge or both edges) are better encoded at birth.

In sum, Experiment 1 revealed that, even from birth, the encoding of syllabic information in a sequence depends on their position in the sequence. Edge positions are more accurately encoded than internal positions. But what are the implications for language processing? Language involves encoding multiple hierarchical elements from a sequential stream— syllables combine to form words, words combine to form phrases, and phrases to form sentences. The positions of the items in each sequence are crucial to retain the meanings across each of these levels. How can the positional information

across an entire sentence be encoded if the positions of the internal elements are poorly encoded? One possibility is that prosodic segmentation cues (e.g., pauses or pitch contours) break up otherwise continuous speech and provide cues to word edges and phrasal boundaries with sequential processing constraints operating across each segment. We address this question in Experiment 2.

Experiment 2

Previous research suggests that prosodic cues can, in fact, signal the hierarchical constituent structure of language (Hawthorne & Gerken, 2014; Jusczyk, 2000; Langus, Marchetto, Bion, & Nespor, 2012; Nespor & Vogel, 2007). For example, both adults and infants expect words to be contained within prosodic constituents delimited by boundaries and infants segment novel words better from the edge of a sentence or phrase than from the middle (Johnson & Jusczyk, 2001; Langus et al., 2012; Seidl & Johnson, 2006; Shukla, White, & Aslin, 2011). Peña and colleagues (Peña et al., 2002) demonstrated that adults were unable to learn an AxC pattern of words from continuous speech but succeeded when a consciously imperceptible 25ms pause was inserted between the words. These findings show that older infants and adults can use prosodic cues to segment continuous speech into discrete elements, and that these cues affect the way sequential information is processed. In Experiment 2, we examined whether newborns use segmentation cues to break up a continuous sequence into smaller, discrete elements, with sequential processing constraints operating across each of the subcomponents.

As a cue for segmentation, we inserted a 25ms silent pause in the middle of the sequence, between the third and fourth syllables (see Table 1). If this cue facilitates segmentation, the six-syllable sequence would be divided into two shorter 3-syllable sequences. The 3rd and 4th syllables– that were previously internal to the long sequence – now became the final syllable of the first segment and the first syllable of the second segment, both in edge positions. If sequential processing constraints operate across each segment, the 3rd and 4th syllables would now be encoded as edges and we should observe an increase in the HbO response when those syllables switch positions.

We tested an additional 16 neonates using a NIRS testing protocol identical to the one used in Experiment 1 (see Fig. 1D), with the only difference being the addition of a 25ms pause during the familiarization and test sequences. The six-syllable familiarization sequence contained a 25ms silent pause between the third and fourth syllables (e.g., simebu_tal*e*fo, see Table 1). In the test blocks the same sequence (including the 25ms pause) was presented, but the third and fourth syllables had switched positions (e.g., sime**ta_bu**l*e*fo). The only difference between this condition (Pause Switch condition) and the Internal Switch condition of Experiment 1 was the addition of the 25ms of silence between the two middle syllables. If neonates' brains responded to the syllable switch, the HbO would increase during test blocks (as in the Edge Switch condition of Experiment 1); if they did not detect the syllable switch, even with the addition of the segmentation cue, the HbO would not increase during the test block (as in the Internal Switch condition of Experiment 1).

Experiment 2: Methods

The methodology used in Experiment 2 was identical to that of Experiment 1 except for the modifications listed below.

Participants

Experiment 2 included 16 neonates (10 females; mean age 2.75 days, range 1-5 days; mean gestational age 39.0 weeks, SD = 1.0 weeks; mean weight 3.267 Kg, SD = 0.361Kg). Additional infants were excluded due to motion artifacts (n = 8), and failure to acquire a good signal due to thick hair (n = 4).

Stimuli

In the Pause Switch condition, a 25ms pause was only inserted between the 3rd and 4th syllables of the familiarization and test sequences using MBROLA (Dutoit et al., 1996). We included the pause in both familiarization and test sequences to ensure that infants responded to the switch in syllables and not to a change in the presence of a pause.

Data Processing

HRF rejection: In the Pause Switch condition, considering all tested infants, a 48% of all the HRFs were automatically rejected, and a 9% manually rejected. Each infant contributed an average of 3.25 test blocks (SD = 1.24). Excluding the posterior channels that contributed less data (channels 8, 11, 22, 23 and 24), the number of infants providing good data per channels was on average 12.53 (SD = 1.47).

Statistical Analyses

Cluster-Based Permutation Analysis. In Experiment 2 the Pause Switch condition was compared to both the Internal Switch and the Edge Switch conditions from

Experiment 1. As in Experiment 1 the data were analyzed both with and without the posterior channels.

Mean Activation Analysis. In order to compare across Experiments 1 and 2 and to provide support that these findings result from differences in the manipulation of the test stimuli, rather than differences between the groups of infants, we conducted an additional analysis comparing the responses to the familiarization blocks and test blocks across the three conditions. We performed a two-ways ANOVA using the mean activation data with factors of Condition (Edge Switch/ Internal Switch/ Pause Switch) and Block Type (Familiarization/ Test).

Experiment 2 Results

Cluster Based Permutation Analysis

We found a significantly greater increase in the HbO signal for the Pause Switch condition compared to the Internal Switch condition. A significant cluster was found in the right hemisphere including channels 14, 16, 17, 18, 19, 20, 21, 22 and 23 ($p_{cluster\_1}$ = 0.002), and another significant cluster was found in the left hemisphere including channels 1, 4, 6, 7, 9 and 10 ($p_{cluster\_2}$ = 0.008) (Fig. 3). The same analysis excluding the posterior channels that contributed less data (channels 8, 11, 22, 23 and 24) yielded similar results (cluster 1 included channels 14, 16, 17, 18, 19, 20 and 21, $p_{cluster\_1}$ = 0.008; and cluster 2 channels 1, 4, 6, 7, 9 and 10, $p_{cluster\_2}$ = 0.017). A cluster analysis comparing the Pause Switch condition to the Edge Switch condition revealed no significantly different clusters ($p > .05$), both with the posterior channels included and excluded.

Mean Activation Analysis

The ANOVA with factors of Condition (Edge Switch/ Internal Switch/ Pause Switch) and Block Type (Familiarization/ Test) (Fig. 4 B) found a main effect of Condition ($F(2, 15) = 9.00$, $p = .0002$), a main effect of Block Type ($F(1, 15) = 24.47$, $p = .000003$), and a significant interaction between Condition and Block Type ($F(2, 15) = 12.05$, $p = .00002$). Post-hoc Tukey-Kramer comparisons revealed that the mean activity in the test block of the Internal Switch condition differed from the Edge Switch condition ($p = 10^{-7}$) and the Pause Switch condition ($p = 0.002$) but did not differ from each other ($p > .05$). The Familiarization blocks for the Edge Switch, Internal Switch and Pause Switch condition do not differ between conditions ($p > .05$). Familiarization and Test blocks are different in the Edge Switch ($p = 3x10^{-7}$) condition and in the Pause Switch ($p = .0314$) condition, but not in the Internal Switch condition ($p > .05$).

Experiment 2: Discussion

Experiment 2 revealed that the neonate brain reacted to a positional switch between two internal syllables in a sequence if those syllables are separated by a 25ms pause. We asked if newborns use prosodic boundaries to break up continuous speech into smaller segments, with sequential processing constraints operating across each of the segments. While neonates' brains failed to differentially respond to a positional switch between the middle syllables of a six-syllable sequence in Experiment 1, they responded to the position switch when a subtle 25ms pause was inserted between the middle syllables, suggesting that the pause facilitated encoding of the positional information of the otherwise internal syllables. We propose that the 25ms pause in Experiment 2 segments the six-syllable sequence into two three-syllabic sequences, and that the sequential processing constraints operate across each sequence individually. With the pause, the 3rd

and 4th syllables were encoded as edges – right and left, respectively - of each smaller segment, and that the switch of those syllabic positions affected the brain responses of the neonates.

As in Experiment 1 the differential pattern of increase in HbO between the Internal Switch condition and the Pause Switch condition emerged bilaterally, and in broad regions comprising mostly temporal and frontal areas. This pattern of results in consistent with previous experiments using habituation and change detection paradigms in infants. The pattern of results between the Pause Switch condition and the Edge Switch condition were similar, with no significant spatial or temporal differences revealed by the cluster analysis between them and a majority of significant channels overlapping in their comparisons to the Internal Switch condition. The results from the mean activation analysis support this conclusion with increases in HbO for the Edge Switch and Pause Switch condition, but no increase for the Internal Switch condition. Despite this broad similarity over temporal-frontal regions, the response in the Edge Switch condition seems a slightly more robust than in the Pause Switch condition. In addition, while the cluster analysis revealed broad overlap between the regions that were different from the Internal Switch condition, some channels in the parietal region were involved in the cluster for only the Edge Switch or the Pause Switch comparison to the Internal Switch. Channel 5, on the left hemisphere, appeared only in the cluster comparing the Edge Switch and Internal Switch while channels 15 and 20 on the right hemisphere appeared only in the cluster comparing the Pause Switch and the Internal switch conditions. However, as the overall comparison revealed no differences between the Edge Switch and Pause Switch

conditions, it is difficult to claim that any specific regions responded differently. It remains a question for future research to investigate the processing differences between the switches at the sequence edges and the internal edges marked by prosodic boundaries in either the localization of the regions responding to the change or the time course of the response.

Experiment 2 also provides additional evidence to support the fact that the results obtained in Experiment 1 can be attributed to precise encoding of edges, rather than the number of syllables between the switched elements. Previous work has suggested that the edges are more precisely encoded by manipulating the length of sequences and by controlling for the number of syllables between the switched syllables (Benavides-Varela & Mehler, 2014; Gupta et al., 2005; Henson, 1998). Our converging evidence demonstrates that the newborn brain responds to a switch between two adjacent internal syllables in identical syllabic sequences when provided with a cue indicating an edge between those syllables. If an edge is inserted between the internal syllables, positional encoding at this boundary is enhanced. This indicates that the edges syllables of sequences are more precisely encoded than internal syllables even if the edge is defined by a subtle cue.

General Discussion

In two neuroimaging experiments we demonstrated that neonates are able to encode sequential information from multisyllabic sequences and that positional information from edge syllables is encoded more efficiently than from internal syllables. From birth, infants also use very brief pauses to segment longer syllabic sequences into smaller sequences, with these positional constraints operating across both smaller

sequences. These results demonstrate that sequential processing biases constrain how linguistic stimuli are encoded from birth. This finding is crucial for our understanding of the mechanisms underlying language acquisition in two key ways.

First, Experiment 1 demonstrated that humans are born with processing constraints that privilege sequence edges; neonates are more sensitive to positional changes of syllables at the edges of sequences than to internal syllables. While previous work had demonstrated an edge processing bias in adults and older infants (Benavides-Varela & Mehler, 2014; Henson, 1998; Hurlstone et al., 2014), it remained unclear whether this bias was the result of a fundamental constraint on sequential processing or a consequence of exposure to sequential processing. For example, language highlights edges with cues such as determiners, morphological regularities, and stress and exposure to these cues may make edges more salient. We demonstrated that even from birth there is an enhanced encoding of sequence edges, indicating that this constraint is an inherent signature of sequential processing. The enhanced encoding of edges in general sequential processing constrains linguistic regularities to appear at the edges of constituents rather than experience with edge-based linguistic regularities causing constraints in general sequential processing.

Second, Experiment 2 revealed that when a segmentation cue appeared in the middle of the sequence, the neonate brain responded to the positional switch between the internal syllables straddling the segmentation cue. From birth, prosodic cues segment long sequences into smaller ones, indicating that sequential processing constraints operate in parallel across these multiple segments. This ability is fundamentally important for language acquisition because language requires encoding of multiple hierarchical

levels (e.g., encoding the order of syllables in words and the order of words in sentences) from a single sequential stream. Our results suggest that the fundamental mechanisms for tracking syllables across different hierarchical levels are evident from birth. Prosodic cues have been found to mark boundaries across different hierarchical levels (Jusczyk, 2000; Langus et al., 2012; Nespor & Vogel, 2007), and we demonstrate that newborns' brains are capable of processing subtle cues to segment and encode longer sequences. Our results also suggest that from birth, prosodic boundaries embedded in the speech signal can be used to track the edges of units in the sequential speech stream, a necessary capacity to encode language across different hierarchies.

The current experiments raise questions for future research regarding the mechanisms underlying these processing constraints and their generality across different domains at birth. Despite decades of research examining serial position effects in adults, the specific mechanisms underlying this edge bias and whether similar mechanisms operate across different domains are still unclear (Gupta et al., 2005; Hurlstone et al., 2014; Lehman & Malmberg, 2013). Researchers have examined the role of a variety of different mechanisms to explain the enhanced encoding of edges, including the role of memory buffers (Atkinson & Shiffrin, 1968; Lehman & Malmberg, 2013), chaining - in which associations are formed between adjacent elements of the sequence (Ebbinghaus, 2013) - statistical learning - in which the probabilities of specific items in a sequence appearing next to each other are computed (e.g., Aslin, Saffran, & Newport, 1998; Saffran, Aslin, & Newport, 1996), and masking - in which the amount of flanking interferes with encoding and edge elements are better encoded since they are flanked by one element while internal elements are flanked by two other elements (Gupta et al.,

2005), among others. While the current results do not allow us to disentangle between the underlying mechanisms, we can conclude that humans are born with a bias to encode the internal components of a sequence less efficiently than edge components. By demonstrating that an edge bias is evident from birth and that newborns are sensitive to subtle experimental manipulations, we show that it is feasible to investigate these questions from a developmental perspective.

Our findings also raise questions about the generalizability of our findings. Although we focus on syllabic sequences, these constraints on sequential processing potentially extend to non-linguistic sequences as well. The edge bias in sequential processing is robust in adults across a range of tasks and sensory domains (Endress & Mehler, 2009; Endress et al., 2009; Fournier, Gallimore, Feiszli, & Logan, 2014; Gupta et al., 2005; Gupta, 2003; Henson, 1998; Hurlstone et al., 2014; Murdock Jr., 1962). Since our current experiments only used sequences of syllables, future research will have to uncover how general sequential processing operates at birth and whether the role of segmentation cues in sequential processing can be generalized. If the sequential encoding mechanism is a general one, then similar results should also emerge with non-linguistic stimuli as well, though potentially with domain-specific segmentation cues. Even within linguistic stimuli, there are open questions about whether other cues beside pauses can facilitate segmentation. Natural language contains cues (e.g., pauses, stress markers, prosodic contours, determiners, transitional probabilities) that mark different hierarchical boundaries (Gervain & Werker, 2013; Greenberg, 1957; Hawthorne & Gerken, 2014; Hirsh-Pasek et al., 1987; Hochmann, 2013; Kager, 1995; Morris Halle & Vergnaud, 1987; Nelson, Hirsh-Pasek, Jusczyk, & Cassidy, 1989; Saffran et al., 1996) and each of

these boundaries could signal an edge. Since we focused our experiment specifically on the brief pause as a boundary cue, it has to be explored whether other cues can similarly facilitate sequence segmentation form birth, and whether serial position effects operate across them.

To conclude, our findings advance our understanding of language processing and acquisition by addressing how neonates encode multisyllabic sequences. Humans are born with specific constraints on encoding multisyllabic sequences but also with the ability to use subtle cues in the speech signal to segment the sequential streams of syllables that make up language. The hierarchical organization of syllabic sequences is a signature of language and our results suggest that the fundamental mechanisms that may track this organization are present from birth.

References

Benavides-Varela, S., Gómez, D. M., Macagno, F., Bion, R. A. H., Peretz, I., & Mehler, J. (2011). Memory in the Neonate Brain. *PLoS ONE*, *6*(11), e27497. doi:10.1371/journal.pone.0027497

Benavides-Varela, S., Hochmann, J.-R., Macagno, F., Nespor, M., & Mehler, J. (2012). Newborn's brain activity signals the origin of word memories. *Proceedings of the National Academy of Sciences*, *109*(44), 17908–17913. doi:10.1073/pnas.1205413109

Benavides-Varela, S., & Mehler, J. (2014). Verbal Positional Memory in 7-Month-Olds. *Child Development*, n/a–n/a. doi:10.1111/cdev.12291

Bortfeld, H., Wruck, E., & Boas, D. A. (2007). Assessing infants' cortical response to speech using near-infrared spectroscopy. *NeuroImage*, *34*(1), 407–415. doi:10.1016/j.neuroimage.2006.08.010

Bryant, G. A., & Barrett, H. C. (2007). Recognizing Intentions in Infant-Directed Speech Evidence for Universals. *Psychological Science*, *18*(8), 746–751. doi:10.1111/j.1467-9280.2007.01970.x

Celsis, P., Boulanouar, K., Doyon, B., Ranjeva, J. P., Berry, I., Nespoulous, J. L., & Chollet, F. (1999). Differential fMRI Responses in the Left Posterior Superior Temporal Gyrus and Left Supramarginal Gyrus to Habituation and Change Detection in Syllables and Tones. *NeuroImage*, *9*(1), 135–144. doi:10.1006/nimg.1998.0389

Cohen, J., MacWhinney, B., Flatt, M., & Provost, J. (1993). PsyScope: An interactive graphic system for designing and controlling experiments in the psychology laboratory using Macintosh computers. *Behavior Research Methods, Instruments, & Computers*, *25*(2), 257–271. doi:10.3758/BF03204507

Dehaene-Lambertz, G., Dehaene, S., Anton, J.-L., Campagne, A., Ciuciu, P., Dehaene, G. P., … Poline, J.-B. (2006). Functional segregation of cortical language areas by sentence repetition. *Human Brain Mapping*, *27*(5), 360–371. doi:10.1002/hbm.20250

Dehaene-Lambertz, G., Hertz-Pannier, L., Dubois, J., Mériaux, S., Roche, A., Sigman, M., & Dehaene, S. (2006). Functional organization of perisylvian activation during presentation of sentences in preverbal infants. *Proceedings of the National Academy of Sciences*, *103*(38), 14240–14245. doi:10.1073/pnas.0606302103

Dutoit, T., Pagel, V., Pierret, N., Bataille, F., & Van der Vrecken, O. (1996). The MBROLA project: towards a set of high quality speech synthesizers free of use for non commercial purposes. In *, Fourth International Conference on Spoken Language, 1996. ICSLP 96. Proceedings* (Vol. 3, pp. 1393–1396 vol.3). doi:10.1109/ICSLP.1996.607874

Ebbinghaus, H. (2013). Memory: A Contribution to Experimental Psychology. *Annals of Neurosciences*, *20*(4). doi:10.5214/ans.0972.7531.200408

Endress, A. D., & Mehler, J. (2009). Primitive computations in speech processing. *The Quarterly Journal of Experimental Psychology*, *62*(11), 2187–2209. doi:10.1080/17470210902783646

Endress, A. D., Nespor, M., & Mehler, J. (2009). Perceptual and memory constraints on language acquisition. *Trends in Cognitive Sciences*, *13*(8), 348–353. doi:10.1016/j.tics.2009.05.005

Fernald, A. (1989). Intonation and Communicative Intent in Mothers' Speech to Infants: Is the Melody the Message? *Child Development*, *60*(6), 1497. doi:10.2307/1130938

Fernald, A. (1992). Human maternal vocalizations to infants as biologically relevant signals: An evolutionary perspective. In J. H. Barkow, L. Cosmides, & J. Tooby (Eds.), *The adapted mind: Evolutionary psychology and the generation of culture* (pp. 391–428). New York, NY, US: Oxford University Press.

Fournier, L. R., Gallimore, J. M., Feiszli, K. R., & Logan, G. D. (2014). On the importance of being first: Serial order effects in the interaction between action plans and ongoing actions. *Psychonomic Bulletin & Review*, *21*(1), 163–169. doi:10.3758/s13423-013-0486-0

Gervain, J., Mehler, J., Werker, J. F., Nelson, C. A., Csibra, G., Lloyd-Fox, S., … Aslin, R. N. (2011). Near-infrared spectroscopy: A report from the McDonnell infant methodology consortium. *Developmental Cognitive Neuroscience*, *1*(1), 22–46. doi:10.1016/j.dcn.2010.07.004

Gervain, J., Nespor, M., Mazuka, R., Horie, R., & Mehler, J. (2008). Bootstrapping word order in prelexical infants: A Japanese–Italian cross-linguistic study. *Cognitive Psychology*, *57*(1), 56–74. doi:10.1016/j.cogpsych.2007.12.001

Gervain, J., & Werker, J. F. (2013). Prosody cues word order in 7-month-old bilingual infants. *Nature Communications*, *4*, 1490. doi:10.1038/ncomms2430

Greenberg, J. H. (1957). *Essays in linguistics* (Vol. vii). Chicago, IL, US: University of Chicago Press.

Gulya, M., Galluccio, L., Wilk, A., & Rovee-Collier, C. (2001). Infants' long-term memory for a serial list: Recognition and reactivation. *Developmental Psychobiology*, *38*(3), 174–185. doi:10.1002/dev.1011

Gulya, M., Rovee-Collier, C., Galluccio, L., & Wilk, A. (1998). Memory Processing of a Serial List by Young Infants. *Psychological Science*, *9*(4), 303–307. doi:10.1111/1467-9280.00060

Gupta, P. (2005). Primacy and recency in nonword repetition. *Memory*, *13*(3-4), 318–324. doi:10.1080/09658210344000350

Gupta, P., Lipinski, J., Abbs, B., & Lin, P.-H. (2005). Serial position effects in nonword repetition. *Journal of Memory and Language*, *53*(1), 141–162. doi:10.1016/j.jml.2004.12.002

Hawthorne, K., & Gerken, L. (2014). From pauses to clauses: Prosody facilitates learning of syntactic constituency. *Cognition*, *133*(2), 420–428. doi:10.1016/j.cognition.2014.07.013

Hayes, B. (1995). *Metrical Stress Theory: Principles and Case Studies*. University of Chicago Press.

Henson, R. N. A. (1998). Short-Term Memory for Serial Order: The Start-End Model. *Cognitive Psychology*, *36*(2), 73–137. doi:10.1006/cogp.1998.0685

Hirsh-Pasek, K., Kemler Nelson, D. G., Jusczyk, P. W., Cassidy, K. W., Druss, B., & Kennedy, L. (1987). Clauses are perceptual units for young infants. *Cognition*, *26*(3), 269–286. doi:10.1016/S0010-0277(87)80002-1

Hochmann, J.-R. (2013). Word frequency, function words and the second gavagai problem. *Cognition*, *128*(1), 13–25. doi:10.1016/j.cognition.2013.02.014

Huppert, T. J., Diamond, S. G., Franceschini, M. A., & Boas, D. A. (2009). HomER: a review of time-series analysis methods for near-infrared spectroscopy of the brain. *Applied Optics*, *48*(10), D280–D298. doi:10.1364/AO.48.00D280

Hurlstone, M. J., Hitch, G. J., & Baddeley, A. D. (2014). Memory for serial order across domains: An overview of the literature and directions for future research. *Psychological Bulletin*, *140*(2), 339–373. doi:10.1037/a0034221

Johnson, E. K., & Jusczyk, P. W. (2001). Word Segmentation by 8-Month-Olds: When Speech Cues Count More Than Statistics. *Journal of Memory and Language*, *44*(4), 548–567. doi:10.1006/jmla.2000.2755

Jusczyk, P. W. (2000). *The Discovery of Spoken Language*. MIT Press.

Kager, R. W. J. (1995). Consequences of catalexis. *HIL Phonology Papers*, *1*(January), 269–298.

Langus, A., Marchetto, E., Bion, R. A. H., & Nespor, M. (2012). Can prosody be used to discover hierarchical structure in continuous speech? *Journal of Memory and Language*, *66*(1), 285–306. doi:10.1016/j.jml.2011.09.004

Lloyd-Fox, S., Blasi, A., & Elwell, C. E. (2010). Illuminating the developing brain: The past, present and future of functional near infrared spectroscopy. *Neuroscience & Biobehavioral Reviews*, *34*(3), 269–284. doi:10.1016/j.neubiorev.2009.07.008

Mahmoudzadeh, M., Dehaene-Lambertz, G., Fournier, M., Kongolo, G., Goudjil, S., Dubois, J., … Wallois, F. (2013). Syllabic discrimination in premature human infants prior to complete formation of cortical layers. *Proceedings of the National Academy of Sciences*, *110*(12), 4846–4851. doi:10.1073/pnas.1212220110

MATLAB and Statistics Toolbox Release 2012b. (n.d.). Natick, MA, USA: The MathWorks, Inc.

Morris Halle, & Vergnaud, J.-R. (1987). *An essay on stress.* Cambridge: MIT Press.

Murdock Jr., B. B. (1962). The serial position effect of free recall. *Journal of Experimental Psychology*, *64*(5), 482–488. doi:10.1037/h0045106

Nakano, T., Watanabe, H., Homae, F., & Taga, G. (2009). Prefrontal Cortical Involvement in Young Infants' Analysis of Novelty. *Cerebral Cortex*, *19*(2), 455–463. doi:10.1093/cercor/bhn096

Nelson, D. G. K., Hirsh-Pasek, K., Jusczyk, P. W., & Cassidy, K. W. (1989). How the prosodic cues in motherese might assist language learning. *Journal of Child Language*, *16*(01), 55–68. doi:10.1017/S030500090001343X

Nespor, M., & Vogel, I. (2007). *Prosodic Phonology: With a New Foreword*. Walter de Gruyter.

Oostenveld, R., Fries, P., Maris, E., & Schoffelen, J.-M. (2010). FieldTrip: Open Source Software for Advanced Analysis of MEG, EEG, and Invasive Electrophysiological Data. *Computational Intelligence and Neuroscience*, *2011*, e156869. doi:10.1155/2011/156869

Peña, M., Bonatti, L. L., Nespor, M., & Mehler, J. (2002). Signal-Driven Computations in Speech Processing. *Science*, *298*(5593), 604–607. doi:10.1126/science.1072901

Peña, M., Maki, A., Kovačić, D., Dehaene-Lambertz, G., Koizumi, H., Bouquet, F., & Mehler, J. (2003). Sounds and silence: An optical topography study of language recognition at birth. *Proceedings of the National Academy of Sciences*, *100*(20), 11702–11705. doi:10.1073/pnas.1934290100

Saffran, J. R., Aslin, R. N., & Newport, E. L. (1996). Statistical Learning by 8-Month-Old Infants. *Science*, *274*(5294), 1926–1928. doi:10.1126/science.274.5294.1926

Seidl, A., & Johnson, E. K. (2006). Infant word segmentation revisited: edge alignment facilitates target extraction. *Developmental Science*, *9*(6), 565–573. doi:10.1111/j.1467-7687.2006.00534.x

Shukla, M., White, K. S., & Aslin, R. N. (2011). Prosody guides the rapid mapping of auditory word forms onto visual objects in 6-mo-old infants. *Proceedings of the National Academy of Sciences*, *108*(15), 6038–6043. doi:10.1073/pnas.1017617108

Sigurd, B., Eeg-Olofsson, M., & Van Weijer, J. (2004). Word length, sentence length and frequency – Zipf revisited. *Studia Linguistica*, *58*(1), 37–52. doi:10.1111/j.0039-3193.2004.00109.x

Thiessen, E. D., Hill, E. A., & Saffran, J. R. (2005). Infant-Directed Speech Facilitates Word Segmentation. *Infancy*, *7*(1), 53–71. doi:10.1207/s15327078in0701_5

Thiessen, E. D., & Saffran, J. R. (2007). Learning to Learn: Infants' Acquisition of Stress-Based Strategies for Word Segmentation. *Language Learning and Development*, *3*(1), 73–100. doi:10.1080/15475440709337001

Edge Switch Condition		Internal Switch Condition		Pause Switch Condition	
Familiarization	Test	Familiarization	Test	Familiarization	Test
simebutalɛ**fo**	**fo**mebutalɛ**si**	sime**buta**lɛfo	sime**tabu**lɛfo	sime**bu_ta**lɛfo	sime**ta_bu**lɛfo
fomebutalɛ**si**	**si**mebutalɛ**fo**	sime**tabu**lɛfo	sime**buta**lɛfo	sime**ta_bu**lɛfo	sime**bu_ta**lɛfo
nɛkalisore**vu**	**vu**kalisore**n**ɛ	nɛka**liso**revu	nɛka**soli**revu	nɛka**li_so**revu	nɛka**so_li**revu
vukalisore**n**ɛ	**n**ɛkalisore**vu**	nɛka**soli**revu	nɛka**liso**revu	nɛka**so_li**revu	nɛka**li_so**revu
gamɛzibeko**tu**	**tu**mɛzibeko**ga**	gamɛ**zibe**kotu	gamɛ**bezi**kotu	gamɛ**zi_be**kotu	gamɛ**be_zi**kotu
tumɛzibeko**ga**	**ga**mɛzibeko**tu**	gamɛ**bezi**kotu	gamɛ**zibe**kotu	gamɛ**be_zi**kotu	gamɛ**zi_be**kotu
ŋelokisubɛ**ma**	**ma**lokisubɛ**ŋe**	ŋelo**kisu**bɛma	ŋelo**suki**bɛma	ŋelo**ki_su**bɛma	ŋelo**su_ki**bɛma
malokisubɛ**ŋe**	**ŋe**lokisubɛ**ma**	ŋelo**suki**bɛma	ŋelo**kisu**bɛma	ŋelo**su_ki**bɛma	ŋelo**ki_su**bɛma
ponivelagu**sɛ**	**sɛ**nivelagu**po**	poni**vela**gusɛ	poni**lave**gusɛ	poni**ve_la**gusɛ	poni**la_ve**gusɛ
sɛnivelagu**po**	**po**nivelagu**sɛ**	poni**lave**gusɛ	poni**vela**gusɛ	poni**la_ve**gusɛ	poni**ve_la**gusɛ
bokɛtaselu**ma**	**ma**kɛtaselu**bo**	bokɛ**tase**luma	bokɛ**seta**luma	bokɛ**ta_se**luma	bokɛ**se_ta**luma
makɛtaselu**bo**	**bo**kɛtaselu**ma**	bokɛ**seta**luma	bokɛ**tase**luma	bokɛ**se_ta**luma	bokɛ**ta_se**luma

Table 1. Familiarization and test sequences for Experiment 1 (Edge Switch and Internal Switch conditions) and for Experiment 2 (Pause Switch condition). Infants were randomly assigned to a condition and to one pair of familiarization and test sequences (e.g., one infant in the Edge Switch condition heard "simebutalɛfo" during each familiarization block and "fomebutalɛsi" during each test block.).

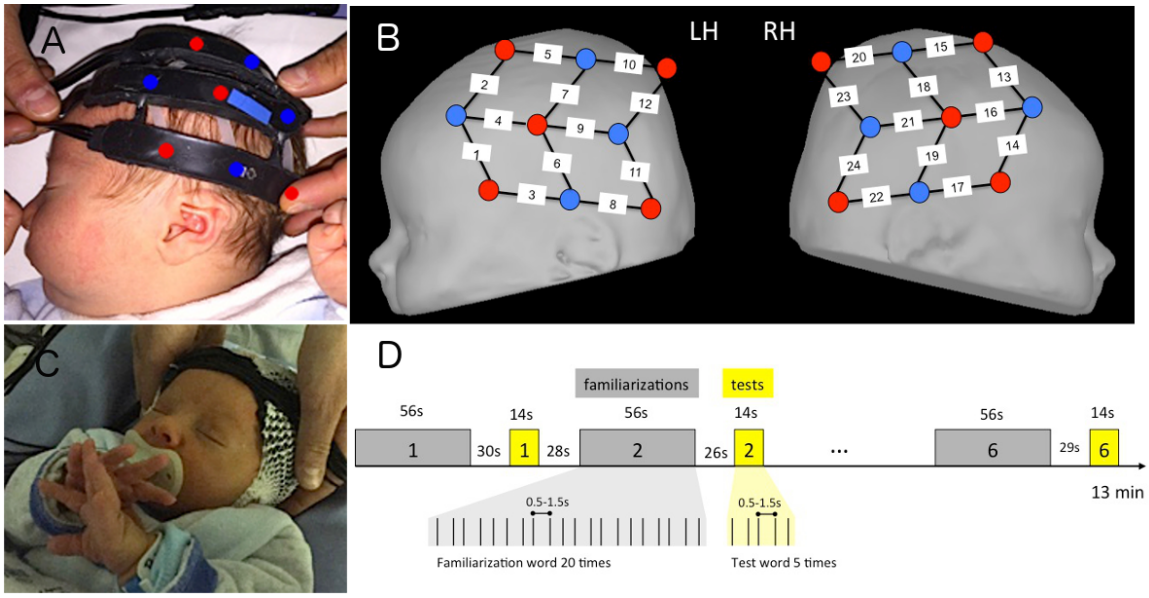

Figure 1. A) The location of one pad showing the location of the channels on an actual infant head. Channels were in the area between light emitters (red dots) and light detectors (blue dots). B) Schematic infant head representing the arrangement of emitters (red dots), detectors (blue dots) and channels (white squares). C) Positioning of the probes during the experiment. D) The experimental design consisted of six alternating blocks of familiarization and test phases, separated by silent recovery periods.

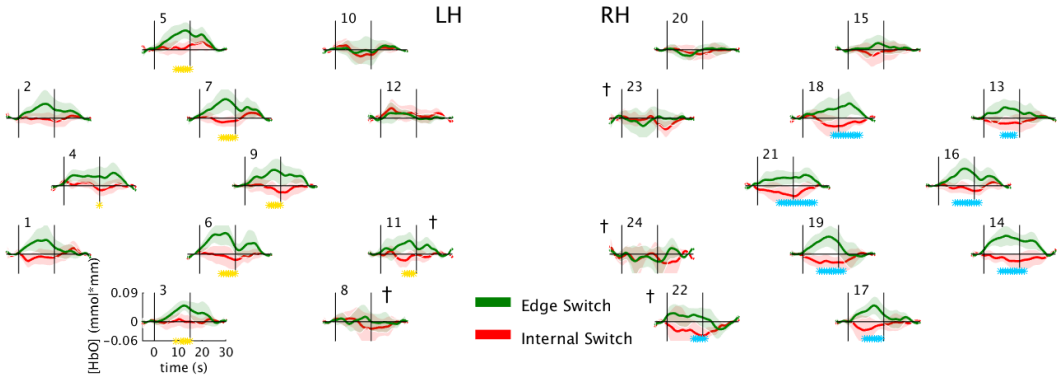

Figure 2. The cluster based analysis results from Experiment 1 (Edge Switch condition compared to Internal Switch condition). Channels are plotted using the same positions as Figure 1a. The x-axis represents time in seconds, the y-axis represents the concentration of HbO in mmol*mm. The vertical lines represent the onset and offset of the test block stimulation. The shaded area is the 95% confidence interval. The yellow and blue asterisks underneath a channel show the samples where the cluster based permutation analysis (1000 randomizations) revealed differences between conditions. Different colored asterisks represent different clusters. Posterior channels found to be more susceptible to exclusion due to a poor signal resulting from inadequate contact with the head are indicated with †.

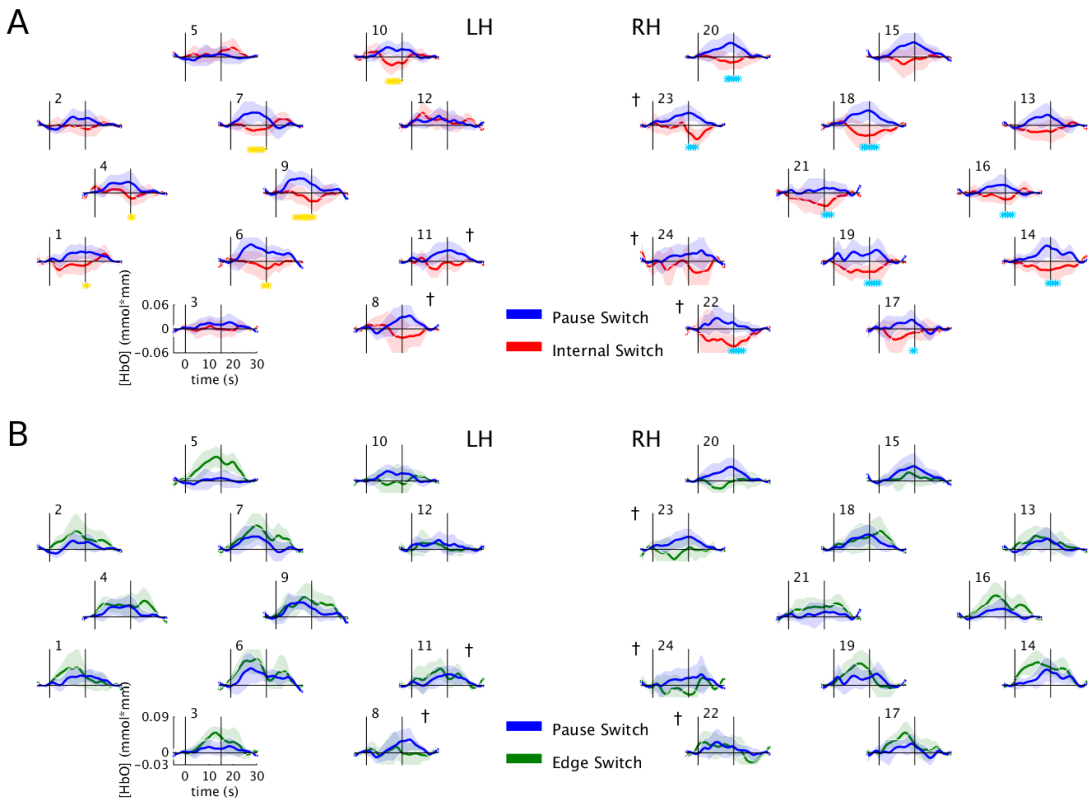

Figure 3. The cluster based analysis results from Experiment 2. A) Pause Switch condition compared to Internal Switch condition. B) Pause Switch condition compared to Edge Switch condition. The x-axis represents time in seconds, the y-axis represents the concentration of HbO in mmol*mm. The vertical lines represent the onset and offset of the test block stimulation. The shaded area is the 95% confidence interval. The yellow and blue asterisks underneath a channel show the samples where the cluster based permutation analysis (1000 randomizations) revealed differences between conditions. Different colored asterisks represent different clusters. Posterior channels found to be more susceptible to exclusion due to a poor signal resulting from inadequate contact with the head are indicated with †.

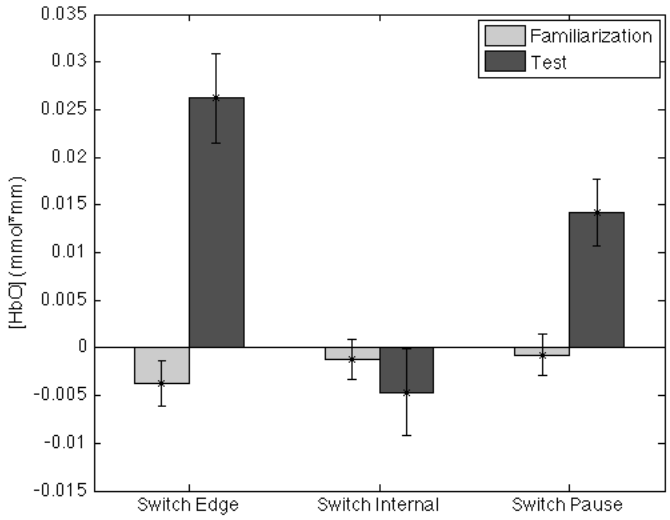

Figure 4. The mean change in HbO for all the channels (excluding 8, 11, 22, 23 and 24) during the test blocks and familiarization blocks in the three conditions. Error bars represent standard errors.

Supplementary Audio File Captions

Supplementary Audio File 1: A sample six-syllable familiarization word used in Experiment 1.

Supplementary Audio File 2: A sample six-syllable Edge-Switch test word used in Experiment 1.

Supplementary Audio File 3: A sample six-syllable Internal-Switch test word used in Experiment 1.

Supplementary Audio File 4: A sample six-syllable familiarization word, containing a 25ms pause between the 3rd and 4th syllables used in Experiment 2.

Supplementary Audio File 4: A sample six-syllable Pause-Switch test word, containing a 25ms pause between the 3rd and 4th syllables used in Experiment 2.