Newborns are sensitive to multiple cues for word segmentation in continuous speech

DOI:
10.1111/desc.12802

Document Version
Accepted author manuscript

Link to publication record in Manchester Research Explorer

Citation for published version (APA):
https://doi.org/10.1111/desc.12802

Published in:
Developmental science

Citing this paper
Please note that where the full-text provided on Manchester Research Explorer is the Author Accepted Manuscript or Proof version this may differ from the final Published version. If citing, it is advised that you check and use the publisher's definitive version.

General rights
Copyright and moral rights for the publications made accessible in the Research Explorer are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

Takedown policy
If you believe that this document breaches copyright please refer to the University of Manchester's Takedown Procedures [http://man.ac.uk/04Y6Bo] or contact uml.scholarlycommunications@manchester.ac.uk providing relevant details, so we can investigate your claim.
Title
Newborns are sensitive to multiple cues for word segmentation in continuous speech

Running Head
Speech segmentation by newborns

Authors
Ana Fló1,2,*, Perrine Brusini1,3, Francesco Macagno4, Marina Nespor1, Jacques Mehler1, and Alissa L. Ferry1,5

1Language, Cognition, and Development Laboratory, Scuola Internazionale di Studi Avanzati, Trieste, Italy.

2Cognitive Neuroimaging Unit, Commissariat à l'Energie Atomique (CEA), Institut National de la Santé et de la Recherche Médicale (INSERM) U992, NeuroSpin Center, 91191 Gif-sur-Yvette, France

3Institute of Psychology Health and Society, University of Liverpool, Liverpool, UK

4Neonatology Unit, Azienda Ospedaliera Santa Maria della Misericordia, Udine, Italy

5Division of Human Communication, Hearing, and Development, University of Manchester, Manchester, UK

*Correspondence to: anaflom@gmail.com

Acknowledgements

We thank the staff at the Neonatology and Obstetrics Departments of the Azienda Ospedaliera Santa Maria della Misericordia for their assistance and the parents of the newborn participants for their participation. We would like to thank Susan Hespos, Yamil Vidal Dos Santos, and Sarà Caviola for feedback and discussions on the manuscript. This research has received funding from the European Research Council under the European Union's Seventh Framework Programme (FP7/2007-2013)/European research Council Grant Agreement 269502 (PASCAL) (to JM).

Competing financial interests

The authors declare not competing financial interests.
Newborns are sensitive to multiple cues for word segmentation in continuous speech

Keywords

Language acquisition, speech segmentation, prosody, statistical learning, newborns, fNIRS.

Research Highlights

• Newborns use statistical distribution cues learned during a familiarization to segment speech and discriminate sequences that complied with those boundaries from sequences that did not.
• Newborns use native language prosodic contours to segment speech and discriminate sequences that complied with those boundaries during a familiarization from sequences that did not.
• When encoding words extracted from continuous speech newborns encode the segmental information (phonology) independently of supra-segmental information (prosody).
• Newborns show bilateral differential responses in a statistical learning segmentation test and right lateralized responses in a prosodic segmentation test.

Abstract

Before infants can learn words, they must identify those words in continuous speech. Yet, the speech signal lacks obvious boundary markers, which poses a potential problem for language acquisition (Swingley, 2009). By the middle of the first year, infants seem to have solved this problem (Bergelson & Swingley, 2012; Jusczyk & Aslin, 1995), but it is unknown if segmentation abilities are present from birth, or if they only emerge after sufficient language
exposure and/or brain maturation. Here, in two independent experiments, we looked at two cues known to be crucial for the segmentation of human speech: the computation of statistical co-occurrences between syllables and the use of the language’s prosody. After a brief familiarization of about 3 minutes with continuous speech, using functional near-infrared spectroscopy (fNIRS), neonates showed differential brain responses on a recognition test to words that violated either the statistical (Experiment 1) or prosodic (Experiment 2) boundaries of the familiarization, compared to words that conformed to those boundaries. Importantly, word recognition in Experiment 2 occurred even in the absence of prosodic information at test, meaning that newborns encoded the phonological content independently of its prosody. These data indicate that humans are born with operational language processing and memory capacities and can use at least two types of cues to segment otherwise continuous speech, a key first step in language acquisition.
1. INTRODUCTION

Before infants can learn what words in their language mean or how they are organized grammatically, they must first segment them from the rapid and continuous stream of sounds that comprise speech. While adults effortlessly extract the words from speech in their native language, speech segmentation poses a potential problem for language acquisition. Speech lacks definitive pauses to mark the word boundaries, raising the question of how infants begin to process this input. Previous studies show that by 6 to 8 months, infants begin to identify words in fluent speech (Bortfeld, Morgan, Golinkoff, & Rathbun, 2005; Jusczyk & Aslin, 1995) and show evidence of understanding the meanings of common nouns (Bergelson & Swingley, 2012), suggesting that infants solve the problem of speech segmentation very early in life. Previous work has focused on two cues present in speech as essential to speech segmentation: the statistical distribution of syllables and prosody. Nevertheless, it remains unknown when infants first use these cues in segmentation: are these mechanisms operational from birth or do they emerge only later in development? It may be that some exposure to language is required for the underlying mechanism to become operative and be able to track the distribution of syllables and/or prosody. Another possibility is that the neonatal brain is too immature to support such a sophisticated language-processing task. Here, we investigate neonates’ capacities to segment and extract words from continuous speech based on the statistical distribution of syllables (Experiment 1) and prosodic information (Experiment 2).

1.1. Statistical Distributional Cues

The fundamental units of speech are phonemes (consonants and vowels) that combine into syllables, which combine into words, which combine into phrases and sentences. Speech however, is not a random collection of phonemes, syllables, or words; instead, each language shows coherent distributional patterns. While some phonemes can co-occur, others cannot
(e.g., phonotactic rules) (Bauer, 2015), and there are sequences of phonemes that are more predictable than others. For example, if an English speaker hears *the bana...*, she is likely to predict that the next syllable is *na*. A more difficult task is to predict which syllable comes after *banana*, as many words can follow *banana* (e.g., split, peel, republic). Because *banana* is a word, these three syllables co-occur, while the syllable coming after *banana* depends on the phrase. The statistical regularities between syllables are one possible cue to word boundaries if the listener assumes that frequently co-occurring syllables form words and that less frequently co-occurring syllables are boundaries between words.

Evidence that older infants and adults can use this strategy to segment speech has been found in a range of experiments using artificial languages (Aslin, Saffran, & Newport, 1998; Saffran, Aslin, & Newport, 1996; Saffran, Newport, & Aslin, 1996). After just minutes of exposure to a stream of four, tri-syllabic words played in random order, with no other cues to the boundaries, infants as young as 8 months (Saffran, Aslin, et al., 1996) can segment the stream and identify the words in a subsequent test phase. When presented with the tri-syllabic unit “words” they can distinguish them from tri-syllabic “part-words”. Part-words are units composed of the final syllables of one word and the first syllables of another word. Thus, these part-words appeared in the speech stream, but crossed this statistical coherence boundary. This suggests a powerful statistical learning mechanism that allows even infants to quickly track the statistical co-occurrences of sounds and find patterns that might indicate word boundaries.

The goal here was to determine if neonates possess this powerful statistical learning mechanism to identify word boundaries. There are several reasons to believe that this might be the case. Evidence of tracking statistical co-occurrences has been found not only with linguistic stimuli, but also with non-linguistic sounds, visual, and even tactile input (Bulf, Johnson, & Valenza, 2011; Conway & Christiansen, 2005; Kirkham et al., 2002; Saffran,
Johnson, Aslin, & Newport, 1999), suggesting that statistical learning may be a broadly general mechanism that does not require a minimum amount of linguistic exposure to operate. Similarly, statistical learning does not appear to be limited to humans; birds and rats use statistical co-occurrences for segmenting sequentially presented information (Santolin & Saffran, 2018; Toro & Trobalón, 2005). Considering the evidence across species and domains, the human brain may be sensitive to the distributional properties of speech from birth. Finally, there is evidence that newborns are at least capable of detecting statistical co-occurrences. Teinonen and colleagues (Teinonen, Fellman, Näätänen, Alku, & Huotilainen, 2009) presented newborns with a modified statistical learning task in which, rather than continuous speech, they presented syllables that were separated by 200 ms for an hour. Using electroencephalogram (EEG) they found differential responses for the first syllables of the conforming triplets relative to the other syllables. Using a similar protocol Kudo and colleagues (Kudo, Nonaka, Mizuno, Mizuno, & Okanoya, 2011) found a differential response for the first tone of a triplet after a shorter familiarization time (8 minutes). The use of pre-segmented syllables and the lack of a post-learning segmentation test leave open the question of whether newborns can actually use these patterns to segment and extract word units from continuous speech. However, these findings do indicate that newborns are at least sensitive to statistical patterns.

However, there are also reasons to suspect that newborns may not be able to use statistical learning to segment words. First, the newborn brain may not be mature enough for complex tasks that require integration across brain regions. Sensory and association regions (Karuza et al., 2013) and the medial temporal lobe and hippocampus (Schapiro, Gregory, Landau, McCloskey, & Turk-Browne, 2014) seem to underlie statistical learning in adults. While newborns have some initial organized brain networks (Doria et al., 2010; Fransson et al., 2009; Fransson, Åden, Blennow, & Lagercrantz, 2011), substantial increases in
myelination occur during the first year, with later development for pathways in associative regions (de Graaf-Peters & Hadders-Algra, 2006; Stiles & Jernigan, 2010), and hippocampal pathways (Gómez & Edgin, 2016; Lavenex & Banta Lavenex, 2013). Some experimental and computational work suggest that the neural correlates of statistical learning may change in development (McNealy, Mazziotta, & Dapretto, 2010; Schapiro, Turk-Browne, Botvinick, & Norman, 2017); thus, newborns may lack the necessary brain resources to engage in integrative tasks like statistical learning. Second, while statistical learning has been demonstrated across domains, ages and species, it is not necessarily the result of a single unique mechanism (Frost, Armstrong, Siegelman, & Christiansen, 2015). Qualitative and quantitative differences in statistical learning may exist due to different learning biases across domains, and changes may occur during development (Newport, 2016). Thus, evidence of statistical learning in other domains or other species does not necessarily imply that newborns have the resources to use it to segment continuous speech.

Experiment 1 will address the question of whether newborns can use statistical co-occurrences to segment continuous speech, and whether they recognize the coherent words in a post-learning phase.

1.2. Prosodic information

The fundamental units of speech are phonemes; however, the same phoneme can vary, even within the same word. Natural speech is not flat; rather, speech has variations in the pitch, intensity and duration creating rhythms, stress patterns, and prosodic contours. These variations are known as the supra-segmental properties of speech, or its prosody. While these do not change the identity of the underlying syllables, they do influence speech perception and processing. Speech is hierarchically organized into prosodically cohesive units from the syllable to phrases and utterances, and several studies demonstrate that this
organization facilitates syntactic processing in adults (Christophe, Gout, Peperkamp, & Morgan, 2003; Christophe, Peperkamp, Pallier, Block, & Mehler, 2004; Millotte, René, Wales, & Christophe, 2008; Nespor & Vogel, 2007; Selkirk, 1984; Shattuck-Hufnagel & Turk, 1996), children (de Carvalho, Lidz, Tieu, Bleam, & Christophe, 2016), and even toddlers (Carvalho, Dautriche, Lin, & Christophe, 2017). Thus, prosody seems to play a relevant role in how speech is segmented and perceived by listeners.

Here we focus on prosody’s potential role in segmentation. Endress and Hauser, (Endress & Hauser, 2010) showed that adults can extract words based on the prosody of unfamiliar languages, indicating that there are some supra-segmental properties, even of unfamiliar languages, that highlight word boundaries. Moreover, many studies show that various prosodic cues can help older children and adults segment individual words from speech (Bagou, Fougeron, & Frauenfelder, 2002; Curtin, Mintz, & Christiansen, 2005; Cutler & Norris, 1988; Gout, Christophe, & Morgan, 2004; Hay & Diehl, 2007; Hayes, 1985; Jusczyk, Cutler, & Redanz, 1993; Jusczyk, Houston, & Newsome, 1999; Yoshida et al., 2010).

While several prosodic cues may be relevant for speech segmentation, we opted to focus on prosodic contours. Using a statistical learning paradigm but with prosodic contours included in the speech stream, Shukla and colleagues (Shukla, Nespor, & Mehler, 2007) showed that adults recognize trisyllabic words that were within a prosodic contour but not trisyllabic words that crossed a contour boundary, suggesting that the prosodic contour facilitated segmentation of continuous speech. Others have shown that infants are sensitive to prosodic contours, discriminating between well-formed (i.e., a single prosodic contour) and ill-formed (i.e., the second half of one prosodic contour and the first half of the subsequent contour) contours (Christophe, Dupoux, Bertoncini, & Mehler, 1994; Gerken, Jusczyk, & Mandel, 1994; Hirsh-Pasek et al., 1987; Mandela & Jusczyka, 1994; Nazzi, Nelson, Jusczyk,
Moreover, infants as young as 6 months better remember coherent units that were previously presented within a well-formed contour (Nazzi et al., 2000) and use prosodic contours to identify words within the speech stream (Shukla, White, & Aslin, 2011). Finally, infant-directed speech, which exaggerates the prosodic contours (Fernald, 1989), facilitated speech segmentation in 7-month-olds, while adult-directed speech in the same design did not (Thiessen, Hill, & Saffran, 2005). These findings suggest that around 6 to 7 months, infants can use prosodic contours to segment words from speech.

Importantly, prosodic contours are potentially a cue that may be available from birth. During the last trimester of pregnancy, the infant auditory system is functional (Moore & Linthicum, 2007) and the infant can hear sounds (in particular the mothers’ speech). From the womb, precise phonological information is lost and only supra-segmental components (e.g., prosodic contours) are available to the infant. At birth, infants prefer their mother’s voice to other voices (DeCasper & Fifer, 1980; Querleu et al., 1984), can discriminate their language from languages of different rhythmic classes (Mehler et al., 1988; Nazzi, Bertoncini, & Mehler, 1998), can recognize the rhythmic patterns of stories read to them during the last trimester (DeCasper & Spence, 1986), and even mimic their native language prosodic patterns in their cries (Mampe, Friederici, Christophe, & Wermke, 2009). Furthermore, subtle prosodic boundaries (e.g., 25ms pauses) influence how neonates encode and recognize sequences of six syllables (Ferry et al., 2016). While not a prosodic contour, it does demonstrate that subtle prosodic cues can shape how newborns encode otherwise continuous speech. In sum, these findings suggest that young infants are sensitive to the prosodic contours of their language. Thus, they are a possible cue to facilitate speech segmentation from birth.
One point to note is that infants must extract the phonological information independent of the underlying prosodic contour. The prosodic contour can serve as a cue to the word boundaries, allowing infants to segment the phonological information but infants should be able to recognize that word independently of that specific prosodic contour. Experiment 2 will test if newborns use the prosodic contours of their language to segment words from otherwise continuous speech and later recognize the phonological information from those segmented words.

1.3. Current Studies

This review suggests that at birth infants can attend to at least some of the information conveyed by the statistical distribution and prosody of speech. Our goal was to investigate if neonates can use these cues to segment fluent speech and extract and encode word-like units. One possibility is that newborns can process phonological distributional properties and prosodic properties of speech and use it for speech segmentation. A second possibility is that some exposure to language is required for one or both of these cues, or that the neonatal brain is still too immature to perform such a task. We address these questions in two independent, but related experiments. In Experiment 1, we investigated speech segmentation when the statistical distribution of syllables was the only cue to word boundaries. In Experiment 2, we investigated speech segmentation when the prosodic contours were the only cue to word boundaries.

Infant speech segmentation research has primarily used looking time or head-turning paradigms (Kemler Nelson et al., 1995). However, these are not well adapted for newborns due to their lack of adequate head and neck control. Thus, we modified the standard speech segmentation paradigm to use functional near-infrared spectroscopy (fNIRS), a non-invasive and silent brain imaging technique, particularly suitable for infants (Aslin, Shukla, &
The neuroimaging paradigm brings two advantages relative to behavioral protocols. First, we can test speech segmentation without requiring behavioral responses, which are difficult to obtain from newborns. Second, the cortical responses could provide insight on the functional architecture of the newborn brain. While fNIRS lacks the incredibly precise localization of fMRI, we can investigate regional and hemispheric differences.

This technique uses near infrared light to measure changes in the concentration of oxygenated hemoglobin (HbO$_2$) and deoxygenated hemoglobin (HbR) in surface cortical regions, allowing for inferences about brain responses to external stimuli. A typical hemodynamic response function (HRF) is an increase in HbO$_2$ and a decrease in HbR. Because HbO$_2$ tends to have a better signal-to-noise ratio than HbR (Gervain et al., 2011), HbO$_2$ responses are primarily reported in the fNIRS literature. However, HbR responses are more comparable with the fMRI BOLD signal (Steinbrink et al., 2006). Here, we report both for data transparency, but we base our conclusion in the better-described and more sensitive measure, HbO$_2$.

We implemented an experimental protocol that capitalizes on the robust habituation and change detection effects observed in fNIRS. It is well-documented that repetition of a 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 (Benavides-Varela et al., 2011; Benavides-Varela, Hochmann, Macagno, Nespor, & Mehler, 2012; Ferry et al., 2016; Mahmoudzadeh et al., 2013; Nakano, Watanabe, Homae, & Taga, 2009) and in adults (Dehaene-Lambertz et al., 2006). In particular, fNIRS habituation experiments with infants often find a relatively broad, bilateral pattern of increased HbO$_2$ response over fronto-temporal regions to a stimulus change after habituation. For example, 3- and 4-month-olds were familiarized to a syllable
(e.g., pa) and showed a broad, bilateral HbO$_2$ increase (particularly in frontal and temporal regions) in response to the presentation of new syllables (e.g., ba), compared to the presentation of the same syllable (Nakano et al., 2009). A similar broad, bilateral, fronto-temporal response was found in neonates when they heard a new word vs a previously presented word (Benavides-Varela et al., 2011, 2012; Ferry et al., 2016). Thus, the idea here was to habituate infants to a speech stream and test their responses to words and to part-words, with part-words expected to lead to a larger bilateral fronto-temporal response.

As in standard speech segmentation paradigms, each experiment consisted of a familiarization and a testing phase. In the familiarization phase, newborns, heard ~3 minutes of continuous speech comprised of four tri-syllabic words. At test, we examined differences in recovery patterns between tri-syllabic words and part-words (e.g., the last two syllables of one word and the first syllable of another word). If infants are sensitive to the cue presented in the learning phase, they should perceive the stream of syllables as a collection of the words, and therefore the isolated words in the test phase should elicit a habituation response while the isolated part-words should elicit a response recovery. A differential response at test indicates that not only were infants sensitive to the statistical or prosodic cue, but also that they were capable of using this cue to segment fluent speech, extract, and subsequently recognize the words during the test phase.

While this design adheres as closely as possible to previous looking time or head-turn paradigms, the use of fNIRS required some key modifications to the paradigm. Because the hemodynamic response is intrinsically slow (on the order of 15 seconds), typical fNIRS experiments use block designs, with silent periods of around 30 seconds between stimulation blocks, for the signal to return to baseline levels. This implies presenting words and part-words in different test blocks, separating them in time, and substantially increasing the study duration. Because the influence of the familiarization may weaken over time, we introduced
short re-familiarization blocks of ~30s before each test block, which consisted of continuous speech with the same properties as the initial long familiarization (Figure 1a).

We recorded hemodynamic responses using two silicone pads, one on each side of the head, with emitters and detectors aligned to form 12 channels per pad (for 24 total channels) (Figure 1b). We expected a robust response spanning across multiple channels over frontal-temporal areas (Ferry et al., 2016; Nakano et al., 2009), thus, we placed the pads trying to maximize the coverage of those areas. The probes on each side of the head mean that we can investigate lateralization effects. Previous work in the change-detection paradigm has shown test phase differential responses bilaterally (Benavides-Varela et al., 2012; Ferry et al., 2016), though with larger responses in the right hemisphere (Nakano et al., 2009). However, the designs have been arguably simpler, with the familiarization phase consisting of the repetition of a single syllable or word and the test phase consisting of that identical token compared to a novel token. Here, the task is more complex; the infant is required to extract information during the familiarization phase and recognize it in a different context at test.

While listening to fluent language appears to engage predominantly the left hemisphere (Dehaene-Lambertz, Dehaene, & Hertz-Pannier, 2002; Fava, Hull, & Bortfeld, 2014; Mahmoudzadeh et al., 2013; Peña et al., 2003; Shultz, Vouloumanos, Bennett, & Pelphrey, 2014), both hemispheres play key roles at different levels of processing. The left hemisphere seems to be more involved in processing stimuli at shorter time scales (phonetic level), while the right hemisphere seems to be more dedicated to the integration of information at longer time scales (prosodic features) (Friederici & Alter, 2004; Poeppel, 2003). A right hemispheric bias for prosody processing has been observed in adults (Friederici & Alter, 2004; Gandour et al., 2004; Meyer, Alter, Friederici, Lohmann, & Von Cramon, 2002; Poeppel, 2003; Zatorre & Belin, 2001), children (Wartenburger et al., 2007) and infants (Homae, Watanabe, Nakano, Asakawa, & Taga, 2006; Homae, Watanabe, Nakano, & Taga, 2007; Wartenburger et al.,
2007). If the processes involved in speech segmentation in Experiments 1 and 2 are lateralized we may expect variations in the lateralized recovery responses at test.

2. EXPERIMENT 1

In Experiment 1, the only cue for speech segmentation was the distribution of syllables. The familiarization streams were built using the same parameters as the original study by Saffran and colleagues (Saffran, Aslin, et al., 1996) that has been well-replicated (Pelucchi, Hay, & Saffran, 2015; Thiessen & Saffran, 2003). We concatenated pseudo-randomly four tri-syllabic words (e.g., lamipe, duvoka, nubefi, telugo), resulting in transitional probabilities (TPs) within words equal to 1, and between words equal to 0.33 (la was always followed by mi and mi was always followed by pe, while pe could be followed either by du, nu, or te).

2.1. Methods

2.1.1. Participants

Participants were healthy, full-term neonates born to Italian-speaking mothers, with an Apgar score ≥ 7 in the first minute and ≥ 8 in the fifth minute, head diameter ≥ 33.0 cm, and no cefalhematomas. Experiment 1 included 40 participants (17 females; mean age 3.3 days; range 2-5 days, except one 7 day-old; mean gestational age 39.1 weeks, range 37-41 weeks; mean weight 3.306 kg, SD 0.416 kg) who provided data without motion artefacts from at least one test block per condition. Participants were recruited from the hospital nursery at Azienda Ospedaliera Santa Maria della Misericordia, in Udine, Italy. Parents provided informed consent. The Ethical Committee of the Scuola Internazionale Superiore di Studi Avanzati approved the study.

2.1.2. Stimuli

Stimuli were synthesized using the it4 Italian female voice of the MBROLA diphone database (Dutoit, Pagel, Pierret, Bataille, & van der Vrecken, 1996). The habituation streams
were continuous with no pauses between syllables, all syllables, both within and across word boundaries, were co-articulated, and all the syllables had a consonant-vowel structure. The speech was flat with all phones having a duration of 150 ms and a constant pitch of 200 Hz. All familiarization streams were ramped up and down during the first and last 6 s, to avoid the start and end of the streams serving as cues.

We built the familiarization streams by concatenating in semi-random order, four tri-syllabic words (Table 1). In order to ensure that results were not driven by low-level properties of the stimuli, two different streams were created by using two sets of words (set A and set B). To synthesize the streams the only restrictions for the concatenation were that the same word could not appear twice in a row, and that two words could not alternate more than three times (the sequence $w_k-w_j-w_k-w_j$, where $w_k$ and $w_j$ are two words, was not allowed). In particular, for the long habituation stream the average transitional probability between words was 0.3319 (SD = 0.0241, range [0.2833, 0.3667]), each word appeared 60 times, and each of the 12 possible part-words 17 to 21 times. For the short habituation streams the average transitional probability between words was 0.3229 (SD = 0.0539, range [0.28, 0.41]), each word appeared 32 times (8 times per stream), and each of the 12 possible part-words between 7 and 12 times. Example audio files are provided as SI.

During the test blocks, either the four words or four part-words (Table 1) were presented separated by silent periods of 0.5 s to 1.5 s. We selected four part-words, two of them of the form $X_1Y_1Y_2$, and two of the form $X_2X_3Y_1$, where $X_i$ and $Y_i$ are syllables of different words. Note that half of the sequences that were part-words in set A were words in set B and vice-versa. Infants were randomly assigned to set A or B. Test blocks containing words and part-words were presented interleaved, and the order of presentation was randomized between participants.
2.1.3. Procedure and data acquisition

Data were recorded using an ETG-4000 NIRS machine (Hitachi Medical Corporation, Tokyo, Japan) with light sources wavelengths of 695 and 830 nm, a 3 cm separation between emitters and detectors, a sampling rate of 10 Hz, and laser power per fiber of 0.75 mW. Each probe consisted of nine fibers, five emitters and four detectors, giving 12 recording points per probe (Figure 1b).

Neonates were tested while lying in their cribs, asleep or in a state of quiet rest, in a dimly lit sound-attenuated booth. Sound stimuli were presented at ≈ 60 dB via two loudspeakers, using PsyScope X software (Cohen, MacWhinney, Flatt, & Provost, 1993). The probes were placed one on each side of the head (right probe: channels 1-12; left probe: channels 13-24). Skull landmarks were used. Specifically, the bottom detector was placed above the ear and the probe was kept aligned along the anterior-posterior direction. The probes were held in place with a bandage and an experimenter holding the infant’s head to ensure that the infant did not turn their head to the side and disrupt the probes during the study. An infrared video camera was used to monitor the infant’s behavior.

2.1.4. Data pre-processing

Data were pre-processed using custom functions and functions of the Homer2 NIRS package (Huppert, Diamond, Franceschini, & Boas, 2009), in MATLAB 2015b, according to a general protocol used previously in other NIRS experiments (for a methodological review see Cooper et al., 2012; Gervain et al., 2011); for some applications see (Fekete, Rubin, Carlson, & Mujica-Parodi, 2011; Ferry et al., 2016; Peña et al., 2003). (1) We identified saturated samples (intensity was greater than 5a.u.) and samples with a low signal to noise (ratio of the power in the frequency bands 0.01-1Hz and 3-10 Hz smaller than 2.5 in a 30 s sliding time window). Channels that showed saturation or low signal to noise ratio during more than 50 %
of the time at any of the two wavelengths were pruned, and later, spatially interpolated (see step 5). (2) We converted intensity to optical density using the Homer2 hmrIntensity2OD function, and we linearly detrended the entire time. (3) We detected motion artefacts using a modified version of the Homer2 hmrMotionArtifactsByChannel function (parameter settings: tMotion = 1s, tMask = ± 0.5s, amp_thres = 1.75 and std_thres = 20). The function was modified to set the threshold independently for each channel and subject, based on the distribution of the maximum absolute change for all the time windows. A segment was considered as affected by motion artifact if fast changes were detected in at least 6 out of the 24 channels at any of the two wavelengths. (4) Motion artefacts were corrected using the Homer2 hmrMotionCorrectPCA by removing 0.97 of the variation of the data (5, 10). (5) We spatially interpolated pruned channels and periods with fast changes in less than 6 channels, using modified Shepard's spatial interpolation with a maximum radius of 4.5 cm. (6) We calculated relative changes in HbO2 and HbR using the Homer2 hmrOD2Conc function (4, 11). (7) We repeated step 3 in order to identify remaining artefacts, and blocks containing artefacts were subsequently excluded from the analysis.

2.1.5. Data rejection

We rejected the HRF for blocks containing un-corrected artefacts (step (7) of the preprocessing). In addition, data was visually inspected, compared to annotations made during the experiment and the video recording, and when outliers were identified, manually rejected. Infants were included in the analysis only if they contributed at least one test block per condition. On average infants provided 1.48 blocks per the word condition, and 1.40 blocks per the part-words condition.

Additional infants were tested but excluded from the final analyses because of excessive motion artefacts (n = 34), failure to complete the experiment due to fussiness (n = 10), poor
signal due to thick hair (n = 8), or technical problems (n = 2). This attrition rate is consistent with other neonate NIRS studies (Lloyd-Fox et al., 2010).

2.1.6. HRFs calculation

Pre-processed data were band-pass filtered between 0.015 and 0.5Hz. HRFs for HbO$_2$ and HbR were extracted for each test block, by taking the data from -5s to +35s from the onset of the stimuli. The responses were baseline corrected for each channel and test block by removing the mean signal from the periods [-5s-0s]. Finally, an average HRF was calculated per infant per channel per condition, and it was used for statistical analysis.

2.1.7. Statistical Analysis

Cluster Based Permutation Analysis

This analysis was initially developed for electroencephalogramme (EEG) research and implemented in the Fieldtrip MATLAB toolbox (Oostenveld, Fries, Maris, & Schoffelen, 2011) and has previously been applied to NIRS data (Edwards, Wagner, Simon, & Hyde, 2015; Ferry et al., 2016; Mahmoudzadeh et al., 2013). The cluster-based permutation analysis is a non-parametric test that identifies significant differences between conditions in spatio-temporal clusters, retaining sufficient statistical power while adequately controlling for the problem of multiple comparisons. We used the Cluster Based Permutation Analysis to compare the HRFs for Test Blocks of words and part-words in the interval [0s, +30s] from the onset (7440 pairs of data points to compare per infant, 24 channels x 310 samples). We performed two-tailed two-sample t-tests and we used 0.05 as threshold p-value to select the pairs of samples to build the clusters. We considered two pairs of samples temporally adjacent if they were consecutive (time difference of 0.1s), and spatially adjacent if they were at a distance < 3 cm (see Supplementary Table 1). We ran 1000 randomizations to obtain the Monte Carlo p-value.
Mean Response Analysis

As a complementary analysis, we also implemented a more commonly used approach in fNIRS studies, which consists of calculating the mean HRF per channel or in certain regions of interest. This analysis generally results in increased spatial sensitivity compared to the cluster analysis, but information about the time course is lost. We calculated the mean response in the period [+10s, +25s] from the onset of the stimuli where we expected the peak in the HRF. We compared the responses for word and part-words in each channel using one sample two tailed t-tests, and corrected P-values using Holm’s correction. Because we located the probes covering mainly the fronto-temporal areas over both hemispheres, where we expected habituation/ novelty responses to the auditory stimulation, and because this response is broad, we did not expect strong differences across channels within each probe. Nevertheless, hemispheric differences may exist. In order to test this, we calculated a mean response per hemisphere per condition, and we performed a repeated measures ANOVA, using condition (words/ part-words) and hemisphere (left / right) as within subject factors.

2.2. Results

The cluster-based permutation analysis for HbO2 showed greater increase for part-words than words in one cluster in the right hemisphere (P = 0.008) and another in the left hemisphere (P = 0.024). Consistently, for HbR, we found a greater decrease for part-words in a cluster of channels in the right hemisphere (P = 0.007), but there was no significance difference in the left hemisphere (P = 0.2168) (see Figure 2a).

Results for the comparison of the average responses to words and part-words per channel are presented in Figure 2b. The repeated measures ANOVA for the average response per hemisphere and per condition revealed, for HbO2, a main effect a condition (F(1,39) = 8.7113, P = 0.0053291, and no significant interaction hemisphere x condition (F(1,39) =
2.5416, P = 0.11896). Instead for HbR we found a main effect a condition (F(1,39) = 5.1499, P = 0.028857), and a marginally significant interaction hemisphere x condition (F(1,39) = 3.9457, P = 0.05405). A post-hoc Tukey test showed that the marginally significant interaction was due to the response towards words and part-words being different in the right hemisphere (P=0.011366) and only marginally different in the left hemisphere (P = 0.092207). This is consistent with the findings of the cluster-based permutation.

2.3. Experiment 1 Discussion

In Experiment 1, we observed a broad differential hemodynamic response for words compared to part-words, with stronger effects over more anterior channels. The differential response was bilateral with no differences across hemispheres. Overall, the HbO₂ response increased more for the part-words than for words at test and the HbR response decreased more for part-words than words. This pattern suggest that neonates showed a habituation response to the words and a change detection response to the part-words. Consistent with previous research (Gervain et al., 2011), the HbO₂ response was more robust than the HbR response, showing a broad bilateral differential response; while the HbR differential response was only marginally significant in the left hemisphere.

The differential response demonstrates that the ability to quickly compute statistical co-occurrences to segment speech, well-documented for older infants, is available at birth. While neonates have very little experience with the segmental properties of speech, the infants’ brain can compute the transitional probabilities between syllables, and extract and encode the segmented words.

The differential response was evident in the anterior channels, consistent with previous studies using similar protocols (Benavides-Varela et al., 2011, 2012; Ferry et al., 2016; Nakano et al., 2009). While precise anatomical information is unavailable, these
anterior channels were placed to cover temporal, frontal and pre-frontal regions. This provides evidence against the classical view of frontal areas being immature and not connected with the language network at birth, and adds to more recent findings showing activity in this region in neonates and pre-terms (Dehaene-Lambertz & Spelke, 2015; Doria et al., 2010; Mahmoudzadeh et al., 2013). While there was no significant hemisphere difference for HbO₂, there was a trend for a stronger differential activation over the right hemisphere, in line with stronger novelty effects in the right hemisphere reported in previous studies (Nakano et al., 2009). A hemispheric difference was found in the HbR component, with larger differences in the right hemisphere.

Overall, Experiment 1 revealed that from birth, infants can segment continuous speech using only statistical distribution cues. This identifies at least one cue that humans are sensitive to at birth that might facilitate speech segmentation. However, statistical distributions may not be the only cue that infants can use. Experiment 2 used the same paradigm to investigate whether neonates can segment speech when the distributional cues are uninformative, and the only cues to segmentation are the prosodic contours.

3. EXPERIMENT 2

In Experiment 2, we built the habituation stream in such a way that the distribution of syllables was not informative for segmentation, and the only cue for word boundaries were prosodic contours. As our primary goal was to determine whether prosodic contours could have any effect on segmentation, we opted to use prosodic contours from the infants’ native language. The four three-syllabic words were concatenated in a fixed order, leading to uniform transitional probabilities of one, and the intonation contours of four Italian phrases where over imposed on the words (Figure 3). While the protocol was the same in both experiments, in that infants had to extract word boundaries from a stream of syllables, there is
an inherent difference between them. In Experiment 1, neonates learned the statistical transitional probabilities during the familiarization and used them for segmentation. In Experiment 2, infants must use the perceptual units created by prosodic contours to segment the words. The prosodic contours were extracted from Italian and thus the participants had pre-natal experience with their structure. Thus, rather than learning the contours during familiarization, infants were applying previously experienced contours to the stream.

Crucially, the test blocks were presented prosodically flat (as in Experiment 1). Therefore, a habituation effect to words and a response recovery to part-words could not be attributable to the detection of well or ill-formed prosodic units, but would indicate that neonates extracted the words delimited by the prosodic contours, and that they encoded the words forming a representation that abstracts over supra-segmental information.

3.1. Methods

3.1.1. Participants

Requirements for recruitment were identical to Experiment 1. Experiment 2 included 40 participants (19 males; mean age 3.2 days; range 1-5 days; mean gestational age 39.2 weeks, rage 37-41 weeks; mean weight 3.400 kg, SD 0.459 kg) who provided data without motion artefacts from at least one test block per condition.

3.1.2. Stimuli

The familiarization streams were built by concatenating four three-syllabic words (Table 1) in a fix order, resulting in uniform TPs of 1. Phonemes had variable duration and pitch, which served as prosodic cues for segmentation. All familiarization blocks were ramped up and down during the first and last 6 s. In order to ensure that results were not driven by low-level properties of the stimuli, we created two different streams by using two sets of words (set A and B), and infants were randomly assigned to familiarization streams.
We created the prosodic contours by recording four Italian three-syllabic sentences from a female native Italian speaker, extracting the pitch and duration of each phoneme, and over-imposing them to the pseudo-words of the stream. To over-impose the pitch and duration, we first fitted the pitch using the `SmoothingSpline` function of MATLAB, and calculated the pitch in 12 equidistant points per phoneme. We normalized the pitch of each phrase to a mean pitch of 200 Hz, and the total duration to 900 ms conserving the relative duration of each phoneme (Figure 3). Finally, we concatenated the contours in a fixed order and used them to synthesize the streams with MBROLA. Note that because both the order of words and the prosodic contours are fixed, the pitch and duration for each phoneme are conserved along the entire familiarization. In other words, each word was marked by a specific prosodic contour. Test items were presented flat, thus were identical to the test blocks for Experiment 1. Example audio files of the stimuli are provided as SI.

3.1.3. **Procedure and data acquisition**

The procedure and data acquisition were identical to Experiment 1.

3.1.4. **Data pre-processing**

The data pre-processing was identical to Experiment 1.

3.1.5. **Data rejection**

The criteria for inclusion was identical to Experiment 1. On average infants provided 1.45 blocks for the Word condition, and 1.52 blocks for the Part-Words condition.

Additional infants were tested but excluded from the final analyses because of excessive motion artefacts (n = 27), failure to complete the experiment due to fussiness (n = 10), poor signal due to thick hair (n = 13), or technical problems (n = 1). This attrition rate is consistent with other studies using NIRS with neonates (Lloyd-Fox et al., 2010) and with Experiment 1.
3.1.6. HRFs calculation

The HRF calculation was identical to Experiment 1.

3.1.7. Statistical Analysis

The statistical analysis was identical to Experiment 1.

3.2. Results

The cluster-based permutation analysis revealed higher HbO$_2$ responses for part-words than for words in a cluster in the right hemisphere ($P = 0.0090$), while the left hemisphere did not show significant differences ($P = 0.2827$ for the strongest cluster candidate). Results for HbR were less conclusive. We found a marginally significant cluster, revealing stronger response for part-words than for words in the right hemisphere in an earlier time window relative to HbO$_2$ ($P = 0.0629$), and no significant results in the left hemisphere (Figure 4a).

Figure 4b shows the Holm corrected $P$-values for the one-sample two-tailed t-tests comparing the mean HbO$_2$ response for words vs part-words. A repeated measure ANOVA on the mean HbO$_2$ response per test condition and hemisphere showed a main effect of condition ($F(1,39) = 7.3442$, $P = 0.0099468$), and a significant interaction hemisphere x condition ($F(1,39) = 6.4311$, $P = 0.015332$). A post-hoc Tukey test revealed that the interaction was due to a higher response for part-words in the right compared to the left hemisphere ($P = 0.0144408$), while the response for words did not differ across hemispheres ($P = 0.28984$). Moreover, the response for words and part-words was significantly different in the right hemisphere ($P = 0.0025402$), but only marginally different in the left hemisphere ($P = 0.063947$). The same analysis with the HbR component did not show any significant effects of the condition and hemisphere.

3.3. Experiment 2 Discussion
We found a differential HbO$_2$ response for words and part-words, with a larger response for part-words. As in Experiment 1, this suggests a habituation response to the words and a change detection response to the part-words. The effect was broad over anterior channels, but restricted to the right hemisphere. While a clear differential response was observed for HbO$_2$, no clear pattern was observed for the HbR component, consistent with previous findings of a better signal-to-noise ratio in HbO$_2$.

The habituation response for words and the recovery response for part-words indicates that infants were more familiar with the words than the part-words. This implies that Italian neonates perceived Italian intonation phrases as perceptual units and used these prosodic contours for word segmentation. Importantly, the prosodic contours were used to identify the word boundaries and extract the identifying phonological information. Words and part-words during the test lacked the prosodic contour and were acoustically different from the familiarization. The fact that we observed differential responses indicates that the newborns extracted and encoded the phonetic content independent of superficial acoustic properties, which is a fundamental requirement for word representations. This is consistent with experiments reporting that young infants (Jusczyk, Pisoni, & Mullennix, 1992; Kuhl, 1983) and even pre-terms (Mahmoudzadeh et al., 2013) can disregard voice differences in phoneme recognition, and it is compatible with a categorical perception of phonemes from birth (Dehaene-Lambertz & Pena, 2001; Eimas & Miller, 1980; Eimas, Siqueland, Jusczyk, & Vigorito, 1971). We build upon these results by showing that newborns can track phonologic information across prosodic variations.

We observed a significant difference between hemispheres, with a response recovery for part-words only in the right hemisphere. This is in contrast to Experiment 1, which showed bilateral differential responses. What could account for these differences? This cannot be due to the specific nature of the test stimuli, which were identical across the two
experiments. We tested the recognition of words vs part-words capitalizing on the strong habituation effects observed in fNIRS (Benavides-Varela et al., 2012; Nakano et al., 2009), but the origin of this effect is not clear so interpretations must be made with caution. One possible explanation is the use of prosody as segmentation cue in Experiment 2. Differentiating words from part-words in Experiment 2 requires integrating supra-segmental auditory information to segment phonological information, while the sequential position of the phonemes is not informative. Thus, if the right hemisphere is more involved in processing prosody, this may explain the response recovery for part-words in the right but not the left hemisphere. An alternative possibility is based on the finding that novelty responses are generally stronger in the right hemisphere. The stimuli in both test blocks of Experiment 2 were modified, in the sense that the prosodic information was removed. This means that both test blocks had some change, which may have minimized the differential processing between the word and part-word test blocks. The current design cannot distinguish between these different possibilities but poses an interesting question for future research.

Finally, we tested infants born to Italian-speaking mothers on their use of Italian prosodic contours to segment speech. The goal was to investigate whether neonates use prosodic contours for segmentation. We show that they can. An open question that this raises though is the role of the prenatal input. Numerous studies demonstrate that even before birth infants actively process auditory stimuli and attend to prosodic information (DeCasper & Fifer, 1980; DeCasper & Spence, 1986; Kisilevsky, Hains, Jacquet, Granier-Deferre, & Lecanuet, 2004; Mampe et al., 2009; Mehler et al., 1988; Nazzi et al., 1998; Querleu et al., 1984). Thus, the infants in our study, born to Italian-speaking mothers, would have had prenatal experience with the prosodic contours of Italian. Since the goal of the current study was whether newborns can use prosodic contours at all in segmentation, we opted to use the familiar contours. However, it is an intriguing question of whether these segmentation skills
are based on the prenatal experience or whether they are more universal. Future research is needed to test whether newborns can segment with unfamiliar prosody to determine whether the sensitivity to prosodic cues is innate or is a consequence of pre-natal experience.

In summary, we show that neonates can use the prosodic contours of their language to segment speech, and that they encode the underlying phonological information independently of the prosodic content.

4. GENERAL DISCUSSION

Overall, our results demonstrate impressive speech segmentation capacities at birth. Newborns can find word boundaries using distributional cues over phonemes in flat speech, as well as prosodic contours when the distribution of phonemes was uninformative. This suggests that newborns have at least two potential sources of information to start segmenting speech.

The current studies were designed to investigate either speech segmentation cue in isolation. While we showed that infants could use both cues here, it opens up several key questions about how infants use these cues in natural language processing. Natural language contains both distributional and prosodic cues while in the current experiments the informative cue was either one or the other. Therefore, we can conclude infants can process both cues, but additional research is necessary to investigate how they are integrated. One possibility is that prosody restricts the time window on which statistical distributional information is integrated, as has been shown in adults (Shukla et al., 2007). Therefore, prosody may help segmentation by restricting the space of possibilities for the phonological content. Additional research on the integration of prosody and distributional cues at birth is needed to investigate this possibility.
To succeed on the current tasks infants needed to not only extract the words from the continuous speech stream, but also recognize and discriminate words and part-words in the test. This indicates that newborns have remarkable short-term memory capacities. While previous work has shown that neonates can recognize familiar words after a short break (Benavides-Varela et al., 2011, 2012; Ferry et al., 2016), the current study expands upon those findings. The previous studies habituated infants to repetitions of isolated words while the current study required infants to extract the words from the continuous stream. Moreover, the previous studies used identical stimuli for the familiarization and familiar test items. Here in Experiment 2, neonates were able to recognize the words based on the phonological content, even when the suprasegmental information changed. Thus, the neonates were able to remember the abstract word form, not just the identical sound.

One notable finding in the study was the presence of a right hemisphere lateralization during the test. While more accentuated in Experiment 2, differential responses were stronger on the right hemisphere in both experiments. This may seem unexpected given that previous neuroimaging work has typically identified a left hemisphere bias in language processing (Dehaene-Lambertz et al., 2002; Peña et al., 2003; Shultz et al., 2014), but it is consistent with previous studies reporting the right hemisphere being more involved in change detection, particularly in areas involved in memory or recognition (Nakano et al., 2009). We believe this is a consequence of the nature of the experimental protocol. Language processing studies with infants typically compare language to silence or a control stimulus (e.g., backwards speech), showing that the left hemisphere is more active when listening to the linguistic stimuli. Here, both test items were words, and thus a general effect of language processing would be unlikely to show substantial differences across the two test conditions. To determine if there are lateralization differences between speech segmentation via
distributional cues or prosodic contours, future research should focus on the activation patterns during the learning period.

Finally, concerning the use of both HbO$_2$ and HbR, previous fNIRS work, particularly with the change detection paradigms has focused primarily on HbO$_2$ with the logic being that the HbR component had a lower signal-to-noise ratio and is thus less informative. Our HbO$_2$ findings were consistent with those reported in previous infant fNIRS change detection paradigms. However, the HbR component should not be ignored completely for several reasons. First, while it is clear that neuronal activity and cerebral blood flow are linked, the specific mechanisms and the roles of HbO$_2$ and HbR components are not clear, even in adults (Steinbrink et al., 2006), and it is premature to rule out the relevance of either component. Second, fMRI only tracks changes in HbR, so the HbR component should be reported to facilitate comparisons across techniques. Finally, fNIRS is a developing technique so we opted to report both results for transparency. Our results were consistent with the fNIRS findings that have reported the HbR component. The HRF for HbR was inverted, generally weaker, and less informative than HbO$_2$. The results were globally consistent with the findings from HbO$_2$, though some effects did not reach significance in both measures, likely as a result from the lower signal-to-noise ratio in the HbR component.

In conclusion, we show that human newborns can segment continuous speech streams by tracking the co-occurrence of syllables and by using prosodic perceptual units to chunk the stream. These findings highlight the impressive mechanisms that are available from birth to start to segment the linguistic input and raise questions about how infants may use these cues with natural speech. The ability to segment words from speech is fundamental to building a vocabulary and identifying the syntactic organization of one’s native language. Our results suggest that at least some of the fundamental mechanisms of speech segmentation are present from birth.
References


https://doi.org/10.3389/fnins.2012.00147

https://doi.org/10.1016/j.cognition.2004.08.005

https://doi.org/10.1037/0096-1523.14.1.113


https://doi.org/10.1016/j.earlhumdev.2005.10.013

https://doi.org/10.1126/science.7375928

https://doi.org/10.1016/0163-6383(86)90025-1


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-

https://doi.org/10.1111/j.1467-7687.2011.01056.x


https://doi.org/10.1016/j.neubiorev.2009.07.008


https://doi.org/10.1073/pnas.1212220110


https://doi.org/10.1016/j.cub.2009.09.064


McNealy, K., Mazziotta, J. C., & Dapretto, M. (2010). The neural basis of speech parsing in


https://doi.org/10.1017/langcog.2016.20

https://doi.org/10.1155/2011/156869

https://doi.org/10.1111/j.1467-8624.2009.01290.x


Poeppel, D. (2003). The analysis of speech in different temporal integration windows: Cerebral lateralization as “asymmetric sampling in time.” In *Speech Communication*


Table 1: Words and part-words used in Experiments 1 and 2.

<table>
<thead>
<tr>
<th>Stream</th>
<th>Words</th>
<th>Part-words</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>lamipe</td>
<td>mipedu</td>
</tr>
<tr>
<td></td>
<td>duvoka</td>
<td>kanube</td>
</tr>
<tr>
<td></td>
<td>nubefi</td>
<td>fitelu</td>
</tr>
<tr>
<td></td>
<td>telugo</td>
<td>lugola</td>
</tr>
<tr>
<td>B</td>
<td>mipedu</td>
<td>peduvo</td>
</tr>
<tr>
<td></td>
<td>vokanu</td>
<td>golami</td>
</tr>
<tr>
<td></td>
<td>befite</td>
<td>nubefi</td>
</tr>
<tr>
<td></td>
<td>lugola</td>
<td>telugo</td>
</tr>
</tbody>
</table>
Figure 1: (a) Protocol for Experiment 1 and Experiment 2. (b) Scheme of the probe locations.
Figure 2: Responses during the Test Blocks for Experiment 1. In (a) HRF towards words and part-words for HbO$_2$ and HbR. The shaded area represents standard errors across subjects. Vertical lines indicate the stimulus onset and offset. Significant clusters of differential responses across time points and channels are indicated by the lines above and below (in red differences in HbO$_2$ and in blue differences in HbR). In (b) t-values for the t-tests comparing the mean differential HbO$_2$ responses for part-words vs. words in the period [+10s, +25s] from the onset of the stimuli. Asterisks indicate channels where significant differences were found (P-values corrected for false discovery rate using Holm’s correction < 0.05.)
Figure 3: Prosodic contours of four Italian phrases over-imposed to the words as segmentation cues. The left column (red) shows the prosodic contours of the real recorded phrases, the right column shows the prosodic contours imposed to the synthesized speech after normalizing the duration and the average pitch (blue).
Figure 4: Responses during the Test Blocks for Experiment 2. In (a) HRF towards words and part-words for HbO$_2$ and HbR. The shaded area represents standard errors across subjects. Vertical lines indicate the stimulus onset and offset. Significant clusters of differential responses across time points and channels are indicated by the lines above and below (in red differences in HbO$_2$ and in blue differences in HbR). In (b) t-values for the t-tests comparing the mean differential HbO$_2$ responses for part-words vs. words in the period [+10s, +25s] from the onset of the stimuli. Asterisks indicate channels where significant differences were found (P-values corrected for false discovery rate using Holm’s correction < 0.05.)
Supplementary Information

Supplementary Table 1: Neighbour channels used for the cluster based permutation tests.

<table>
<thead>
<tr>
<th>Left Probe</th>
<th>Right Probe</th>
</tr>
</thead>
<tbody>
<tr>
<td>Channel</td>
<td>Neighbours</td>
</tr>
<tr>
<td>1</td>
<td>2 3 4</td>
</tr>
<tr>
<td>2</td>
<td>1 4 5</td>
</tr>
<tr>
<td>3</td>
<td>1 6</td>
</tr>
<tr>
<td>4</td>
<td>1 2 6 7</td>
</tr>
<tr>
<td>5</td>
<td>2 7</td>
</tr>
<tr>
<td>6</td>
<td>3 4 7 8 9</td>
</tr>
<tr>
<td>7</td>
<td>4 5 6 9 10</td>
</tr>
<tr>
<td>8</td>
<td>6 11</td>
</tr>
<tr>
<td>9</td>
<td>6 7 11 12</td>
</tr>
<tr>
<td>10</td>
<td>7 12</td>
</tr>
<tr>
<td>11</td>
<td>8 9 12</td>
</tr>
<tr>
<td>12</td>
<td>9 10 11</td>
</tr>
</tbody>
</table>

Supplementary Audio 1. Long familiarization of Experiment 1.

Supplementary Audio 2. Long familiarization of Experiment 2.

Supplementary Audio 3. Words Test Block example.

Supplementary Audio 3. Part-Words Test Block example.