

Adult-like processing of time-compressed speech by newborns: A NIRS study

Cécile Issard^{a,b}, Judit Gervain^{a,b,*}

^a Laboratoire Psychologie de la Perception, Université Paris Descartes, 75006 Paris, France

^b Laboratoire Psychologie de la Perception, Centre National de la Recherche Scientifique UMR 8242, 75006 Paris, France

ARTICLE INFO

Article history:

Received 4 April 2016

Received in revised form 5 October 2016

Accepted 6 October 2016

Available online 29 October 2016

Keywords:

Time-compressed speech

Newborn infants

Near-infrared spectroscopy (NIRS)

Prosody

Temporal envelope

ABSTRACT

Humans can adapt to a wide range of variations in the speech signal, maintaining an invariant representation of the linguistic information it contains. Among them, adaptation to rapid or time-compressed speech has been well studied in adults, but the developmental origin of this capacity remains unknown. Does this ability depend on experience with speech (if yes, as heard in utero or as heard postnatally), with sounds in general or is it experience-independent? Using near-infrared spectroscopy, we show that the newborn brain can discriminate between three different compression rates: normal, i.e. 100% of the original duration, moderately compressed, i.e. 60% of original duration and highly compressed, i.e. 30% of original duration. Even more interestingly, responses to normal and moderately compressed speech are similar, showing a canonical hemodynamic response in the left temporoparietal, right frontal and right temporal cortex, while responses to highly compressed speech are inverted, showing a decrease in oxy-hemoglobin concentration. These results mirror those found in adults, who readily adapt to moderately compressed, but not to highly compressed speech, showing that adaptation to time-compressed speech requires little or no experience with speech, and happens at an auditory, and not at a more abstract linguistic level.

© 2016 The Authors. Published by Elsevier Ltd. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

1. Introduction

Understanding speech is remarkably constant: despite considerable differences in voice quality, accent or speech rates between speakers, we have the subjective impression of hearing the same speech sounds and words under a wide variety of circumstances. Indeed, our auditory and linguistic systems readily normalize the highly variable speech signal in order to extract linguistic units that are necessary for comprehension. One important dimension along which speech may vary considerably is time. Speech rate differs within and across speakers, but this typically doesn't impede communication.

Indeed, successful adaptation to time-compressed speech has been observed in adults and older children, in tasks such as word comprehension (Dupoux and Green, 1997; Orchik and Oelschlaeger, 1977), sentence comprehension (e.g. Ahissar et al., 2001; Peelle et al., 2004) or reporting syllables (Mehler et al., 1993; Pallier et al., 1998; Sebastián-Gallés et al., 2000). In these latter

studies, participants had to report the words or syllables they perceived in sentences compressed to 38%–40% of their initial duration, after having been habituated with compressed speech or after having received no habituation. Specifically, using 40% compression, Mehler et al. (1993) presented French or English sentences to French or English monolinguals and to French-English bilinguals. Participants reported higher numbers of words when they were initially habituated to and then tested in their native language. A follow-up study showed that Spanish-Catalan bilinguals adapted to Spanish or Catalan sentences compressed at 38% after habituation in the other language (i.e. habituation in Spanish before test in Catalan, and habituation in Catalan before test in Spanish). In two subsequent studies, English monolingual adults tested with 40% compressed English sentences benefited from habituation to time-compressed speech in Dutch, which is rhythmically similar to English, and Spanish monolinguals tested with 38% compressed Spanish sentences benefited from habituation with Catalan and Greek, languages that shares rhythmic properties with Spanish (Pallier et al., 1998; Sebastián-Gallés et al., 2000). These studies thus show that adults are able to adapt to moderately compressed speech in their native language, as well as in unfamiliar languages, if those belong to the same rhythmic class as their native language. This suggests that adaption to time-compressed speech is

* Corresponding author at: Laboratoire Psychologie de la Perception, 45 rue des Saints-Pères, 75006, Paris, France.

E-mail address: judit.gervain@parisdescartes.fr (J. Gervain).

based on auditory/phonological mechanisms ('sound-based' adaptation), rather than on top-down linguistic knowledge regarding the lexicon, the syntax or semantics of the native language ('lexical/grammatical' adaptation).

Neuroimaging studies also provided evidence for a dissociation between lexical/grammatical processing and sound-based adaptation to time compressed speech, the two processes involving different neural pathways. In an fMRI study, Peelle et al. (2004) presented syntactically simple or complex sentences compressed to 80%, 65%, or 50% of their normal duration. Time-compressed sentences produced activation in the anterior cingulate, the striatum, the premotor cortex, and portions of temporal cortex, regardless of syntactic complexity. Others studies found that some brain regions, e.g. the Heschl's gyrus (Nourski et al., 2009) and the neighboring sectors of the superior temporal gyrus (Vagharchakian et al., 2012), showed a pattern of activity that followed the temporal envelop of compressed speech, even when linguistic comprehension broke down, e.g. at 20% compression rate. Other brain areas, such as the anterior part of the superior temporal sulcus, by contrast, showed a constant response, not locked to the compression rate of the speech signal for levels of compression that were intelligible (40%, 60%, 80% and 100% compression), but ceased to respond for compression levels that were no longer understandable, i.e. 20% (Vagharchakian et al., 2012).

These studies investigated adults and older children, i.e. participants who are proficient speakers of a language and have considerable experience with speech and language processing in general. Thus, the developmental origins and the existence of a critical period for this ability remain unexplored. It is unclear if adaptation to time-compressed speech can occur independently of any experience with speech (at least with broadcast speech, transmitted through the air, as experienced postnatally). Several hypotheses may be considered. First, this ability might rely on top-down linguistic knowledge of the lexicon and/or the grammar (morphology, syntax, semantics), which helps listeners discover linguistically relevant constants in the time-altered speech stream. In this case, newborns and young infants should fail to adapt to time-compressed speech. This hypothesis is relatively unlikely, since adults and older children can adapt to compressed speech in an unknown language, as long as that is rhythmically similar to their mother tongue. Second, one may assume that the adaptation ability depends on experience with spoken language in general (not specifically with the native language). Such a hypothesis predicts that adults and children can adapt to time-compressed speech, as has been observed, but that newborns, who only have experience with speech as heard in utero, which is very different from regular speech transmitted through the air, would fail. Third, it may be the case that little or no experience is needed for the adaptation ability to occur, so the degraded, low-pass filtered speech signal experienced prenatally, which only preserves the prosodic properties of the native language (Gerhardt et al., 1992; Querleu et al., 1988), may be sufficient. In this case, newborns may adapt to time-compressed speech successfully. Here, we show that this is indeed the case, suggesting that adaptation occurs at the auditory/phonological and not at the lexical/grammatical levels. This finding brings the first developmental evidence to the hypothesis that adaptation to time-compressed speech is an auditory/phonological phenomenon.

Specifically, using near-infrared spectroscopy (NIRS) we tested the ability of the newborn brain to discriminate and to adapt to speech at three levels of compression: (i) normal, uncompressed speech, i.e. 100% of its original duration, (ii) speech compressed at a level comprehensible for adults, i.e. 60% of its initial duration, and (iii) speech compressed at a level that is no longer comprehensible for adults, i.e. 30% of its original duration. NIRS is a powerful and easy-to-use neuroimaging method, well suited to test young infants (Gervain et al., 2011; Rossi et al., 2012). It uses the absorbance

properties of oxygenated and deoxygenated hemoglobin to assess the metabolic correlates of brain activity, i.e. the hemodynamic response function (HRF).

2. Materials and methods

2.1. Participants

Fifty-nine healthy full-term neonates participated in the experiment (mean age 2.34 days, range 1–4 days; 34 females). They were recruited during their stay at the hospital after birth. To participate, newborns had to be full-term (gestational weeks ≥ 37), weight more than 2700 g, and have an Apgar score superior or equal to 8 at 10 min after birth. Newborns' hearing was assessed by a measurement of their oto-acoustic emissions during their stay at the maternity and through screening by a local pediatrician: no hearing disabilities, neurological disorders, prenatal or perinatal complications were reported for any of the participants. A questionnaire filled out by the parents as assessed for parental handedness, antecedents of language or hearing impairments in the family. One parent reported a non-hereditary auditory impairment (mother having a deaf ear following an ear infection). No other relevant condition was present.

Four participants were excluded from the analysis due to crying ($n = 3$) and technical problems ($n = 1$). Data from remaining fifty-five newborns were pre-processed, but thirty-four were not retained for the final analysis due to poor data quality (mostly because of movement artifacts and noise related to dark thick hair). To be retained for the final analysis, participants had to have at least 50% good data in at least two of the three conditions. Importantly, this uniform rejection criterion was applied in batch to all infants whose data was pre-processed, leading to the above reported rejection, prior to statistical analysis. Twenty-one participants were thus included in the final analysis.

All parents gave informed written consent. The study was approved by the local ethics committee at Paris Descartes University (CERES number 2011-013).

2.2. Material

The stimuli were 120 utterances of 11 or 12 syllables, randomly chosen from the CHILDES (MacWhinney, 2000) French corpora and assessed for grammaticality and naturalness by a native French speaker. A native female French speaker recorded the selected utterances in an infant-directed manner. To generate time-compressed speech, we compressed the original utterances to 60% and 30% of their initial duration. The two compression rates, 60% and 30%, were chosen because adults perceive them differently. Indeed, speech compressed at 60% is intelligible for adults, while speech compressed at 30% is not (Pallier et al., 1998). For compression, we used the Audacity software, which allows for the correction of the pitch shift due to compression, maintaining the pitch at the same level as in the original recordings. The compression algorithm (implemented by the "change tempo" function of Audacity) maintained the spectral content of the normal utterances but accelerated the amplitude modulations by a factor of 1.67 for the 60% compression rate and by 3.33 for the 30% compression rate (Fig. 1). Intensity was normalized across stimuli. The utterances were finally concatenated into blocks (see below). Within blocks, utterances were separated by silences of 0.5–1.2 s.

The experiment used an alternating/non-alternating design, often used in behavioral studies to test fine-grained perceptual discrimination (Best and Jones, 1998; Gervain et al., 2014), and successfully implemented with NIRS, (e.g. in Sato et al., 2010; Gervain et al., 2012). Two types of blocks were used (Fig. 2): alternating blocks contained an alternation of normal and accel-

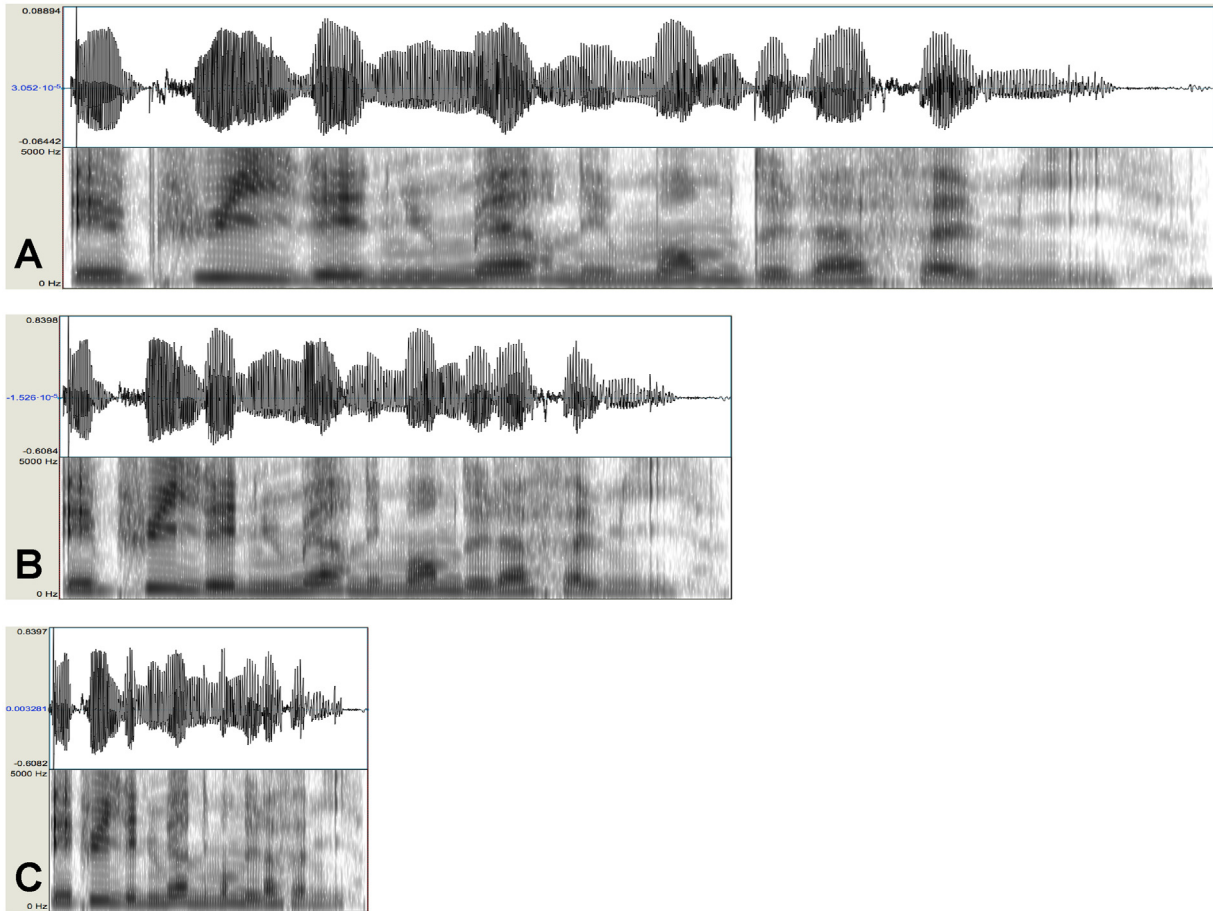


Fig. 1. Waveforms and spectrograms of one sentence at its normal duration (A), compressed to 60% (B) and to 30% (C) of its original duration.

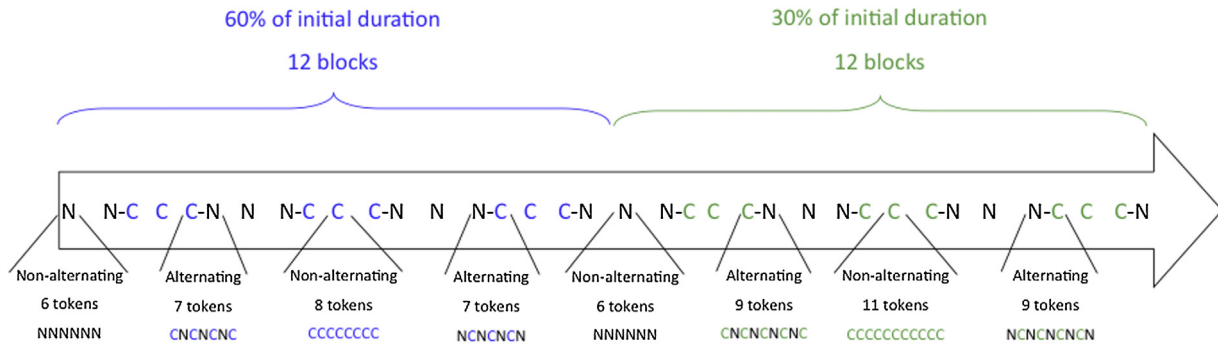


Fig. 2. Experimental design. N: Normal sentences. C: Compressed sentences. All the blocks had the same duration.

erated speech tokens (e.g. 100% and 60% or 100% and 30%), whereas non-alternating blocks contained only one type of speech sounds (only normal speech utterances, only 60% compressed utterances, or only 30% compressed speech). The 30% and 60% compression rates were used in two distinct halves of the experiment (Fig. 2). The order of the 30% and 60% compression halves was counterbalanced across participants. The advantage of the alternating/non-alternating design, e.g. over presenting only non-alternating blocks, is that it can address two different questions. When analyzing non-alternating blocks only, we can explore how and where speech at different compression rates is processed in the newborn brain. Additionally, comparing alternating and non-alternating blocks tells us whether newborns can discriminate between the three different levels of compression.

The number of utterances per block was adjusted such that all five block types (non-alternating 100%, non-alternating 60%, non-alternating 30%, alternating 100% & 60% and alternating 100% & 30%) had approximately the same length (mean duration 18,39 s, range 17–19 s). This choice was made due to the sensitivity of the hemodynamic response function to the length and absolute amount of stimulation. Non-alternating normal blocks thus contained six utterances, 60% compressed non-alternating blocks contained 8 utterances, and 30% compressed non-alternating blocks contained 11 utterances. Blocks alternating between normal and 60% compressed tokens contained 7 utterances and those alternating between normal and 30% compressed tokens contained 9 utterances. Each utterance was presented once per compression rate. The order of the stimuli was randomized within each condi-

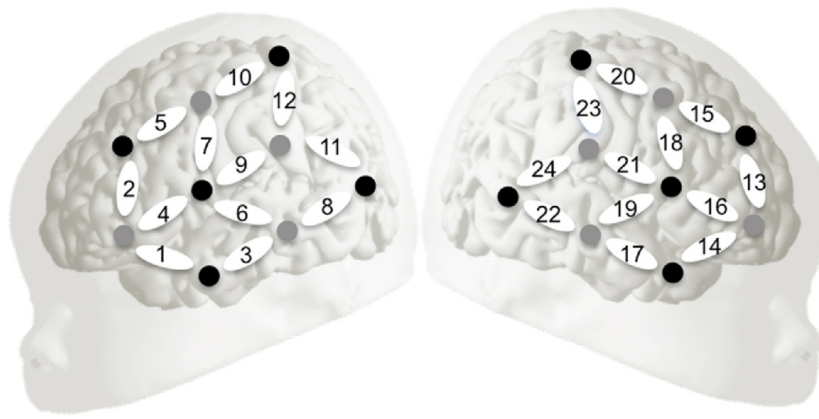


Fig. 3. Optode placement overlaid on the schematic newborn head. Dark circles indicate detectors, light circles indicate sources and numbered ovals correspond to measurement channels.

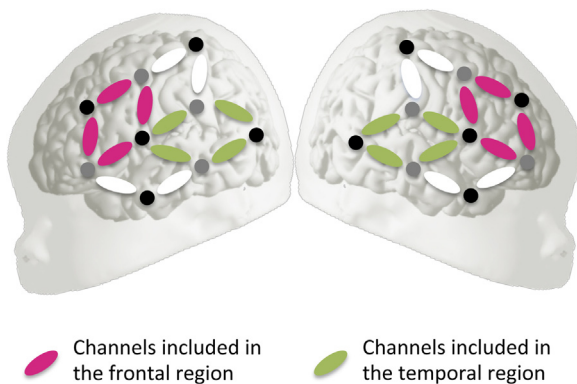


Fig. 4. Channels included in the two regions of interest. Dark circles indicate detectors, light circles indicate sources.

tion. Blocks were separated by pauses varying in duration between 26 and 35 s.

2.3. Procedure

The hemodynamic response of the newborn brain to the auditory stimuli was recorded using a NIRx NIRScout 816 (NIRx Medizintechnik GmbH, Berlin, Germany) NIRS machine (two pulsed wavelengths of 760 nm and 850 nm). LED light sources and detectors were placed on a stretchy cap, each adjacent source-detector pair forming a channel (source–detector separation: 3 cm). The relative concentration of oxy- and deoxyhemoglobin was computed in each channel based on the difference between the intensities of the incident light projected onto the head and the light measured by the detectors, using a modified Beer-Lambert law (for further details, see Gervain et al., 2011). In the apparatus we used, the optodes were placed on the fronto-temporal, temporal and temporo-parietal regions (Fig. 3). This localization was verified against an average newborn MRI head template (Shi et al., 2011) and was already used successfully in previous work from our laboratory (Abboub et al., 2016).

Infants were tested in a quiet room of the hospital, lying in their cribs throughout the 22–25-min testing session, assisted by an experimenter. Parents attended the session. Testing was done with the infants in a state of quiet rest or sleep. The stimuli were presented through two loudspeakers positioned at a distance of approximately 1 m and an angle of 30° from the infant's head respectively. The stimulus presentation was controlled by E-Prime

2.10. The experiment was discontinued if the infant started to cry or upon parental request.

The experiment was divided into two parts: in half of the experiment, we used 60% as the compression rate, in the other half 30%. Each of these two parts contained 12 blocks. The order of the two parts was counterbalanced between participants. The alternating and non-alternating blocks strictly alternated, with half of the infants hearing an alternating block first and the other half hearing a non-alternating block first. If the first non-alternating block was a normal one, the second one was a time-compressed one and vice-versa. Likewise if the first alternating block began with a normal utterance, the second one began with a time-compressed utterance (Fig. 2).

2.4. Data processing and analysis

Data were processed using Matlab (Mathworks) custom scripts developed by the McDonnell consortium “Infant Methodology” (Gervain et al., 2011). First, attenuation changes were converted into changes of concentration in oxy- and deoxyhemoglobin. Data were then filtered to remove components of the signal due to heart-beat, as well as to remove noise, general trends and systemic blood flow variations, using a band pass filter between 0.01 Hz and 1 Hz.

Artifacts (mainly due to movements) were then removed according to the following criteria: each block-channel pair containing a concentration change over $0.1 \text{ mmol} \times \text{mm}$ on two consecutive samples (i.e. 200 ms) was removed from the analysis. Only participants who had at least 50% artifact-free blocks were kept (see Section 2.1 Participants above).

We then computed the mean concentration change in a time window starting 7 s after stimulus onset (to allow for adaptation to take place), and lasting 25 s (comprising the stimulus block starting 7 s after its onset as well as a post-stimulus period of 14 s) in each channel for oxy- and deoxyhemoglobin in each experimental condition. All analyses were carried out for both hemoglobin species, but only oxyhemoglobin yielded significant results, as is common with infant studies (see Gervain et al., 2011 for a review). We thus only report statistical tests for oxyhemoglobin.

We first compared each condition to a zero baseline with permutation tests (Maris and Oostenveld, 2007; Nichols and Holmes, 2002). Permutation tests have the advantage of controlling for the multiple comparisons problem without loss of statistical power, which typically occurs when Bonferroni or other corrections are applied to infant NIRS data (Nichols and Holmes, 2002). The experimental conditions were then compared directly by running permutation tests. Alternating and non-alternating blocks were compared for each compression rate (60% and 30%) separately.

Table 1
Statistical comparisons of the three different types of non-alternating blocks to baseline.

	results
normal	n.s.
60% compression rate	n.s.
30% compression rate	Ch. 10: $p < 0.001$ Ch. 12: $p = 0.025$ Ch. 15: $p < 0.001$ Ch. 22: $p < 0.001$

Table 2
Statistical comparisons of the three types of non-alternating blocks to each other.

main effect of compression rate		Ch. 10: $p = 0.002$
pairwise comparisons	normal vs. 60%	n.s.
	normal vs. 30%	Ch. 10 $p = 0.001$ Ch. 12 $p = 0.020$ Ch. 15 $p = 0.002$ Ch. 22 $p = 0.007$
	60% vs. 30%	Ch. 10: $p = 0.011$

We further analyzed our results using analyses of variance (ANOVAs) looking at specific regions of interest (ROIs). We defined two ROIs per hemisphere, following previous NIRS studies on newborns (Gervain et al., 2012, 2008), as well as the pattern of activity found in response to time-compressed speech in previous imaging studies (Adank and Devlin, 2010; Vagharchakian et al., 2012). We thus defined a temporal region comprising channels 6, 8, 9 and 11 in the left hemisphere and channels 19, 21, 22 and 24 in the right hemisphere, and a frontal region comprising channels 2, 4, 5 and 7 in the left hemisphere and channels 13, 15, 16 and 18 in the right hemisphere. These regions are plotted in Fig. 4.

3. Results

3.1. Channel-by-channel comparisons

3.1.1. Non-alternating blocks

We first analyzed the non-alternating blocks alone. The obtained grand average responses are shown in Fig. 5. The 30% compression non-alternating blocks evoked a significant decrease in oxyhemoglobin as compared to baseline in channels 10, 12, 15, and 22 ($p_{10} < 0.001$, $p_{12} = 0.025$, $p_{15} < 0.001$, $p_{22} < 0.001$, Fig. 6A). No significant channel-by-channel results were obtained for the other two compression rates. The results are summarized in Table 1 and Fig. 6A (only the significant comparisons are shown).

Directly comparing the three compression rates in channel-by-channel permutation tests, we observed a significant effect of compression rate (normal/60%/30%) in channel 10 ($p = 0.002$). Pairwise comparisons with permutations to explore this effect yielded a significant difference between 60% and 30% compressed speech in channel 10 ($p = 0.011$), and between normal and 30% compressed speech in channels 10 ($p = 0.001$), 12 ($p = 0.020$), 15 ($p = 0.002$), and 22 ($p = 0.007$), but no significant difference between normal and 60% compressed speech. These results are summarized in Table 2 and in Fig. 6B–C.

3.1.2. Comparison between alternating and non-alternating blocks

We also compared the hemodynamic response evoked by alternating and non-alternating blocks to the baseline and to each other separately in each channel for each of the compression rates.

For the 60% compression rate, there was a significant increase in oxyhemoglobin concentrations in channel 24 ($p = 0.0396$) for the alternating blocks as compared to baseline. For the 30% compression

Table 3
Statistical comparisons for the alternating and non-alternating (normal and accelerated, either 30% or 60%, averaged together) blocks to baseline and to each other.

alternating blocks vs. baseline	60%	Ch. 24: $p = 0.0396$
	30%	n.s.
non-alternating (normal & accelerated) vs. baseline	60%	n.s.
	30%	n.s.
alternating vs. non-alternating blocks	60%	n.s.
	30%	n.s.

level, we didn't obtain any significant difference between the two.

We didn't obtain any significant difference between alternating and non-alternating blocks in channel-by-channel comparisons for either compression rate.

The above results are summarized in Table 3.

3.2. Analyses of variance

In addition to the above channel-by-channel comparisons, we also conducted analyses of variance (ANOVA) to reveal general patterns in two regions of interests, temporal and frontal, defined on the basis of previous results on the perception of time-compressed speech (Fig. 4).

3.2.1. Non-alternating blocks

An ANOVA with factors Condition (normal/60%/30%), Hemisphere (LH/RH), and ROI (temporal/frontal) as within-subject factors over the three types of non-alternating blocks did not revealed any significant effects or interactions.

3.2.2. Alternating vs. non-alternating blocks

For the 60% compressed part of the experiment, an ANOVA with Block Type (alternating/non-alternating), ROI (frontal/temporal) and Hemisphere (LH/RH) as within-subject factors revealed a main effect of ROI ($F(1, 20) = 7.719$, $p = 0.014$), and an interaction between Block Type and ROI ($F(1, 20) = 8.069$, $p = 0.010$) Fig. 7. As depicted in Fig. 8, the main effect of ROI was due to a higher activity in the temporal regions ($d_{\text{temporal-frontal}} = 1.18 \times 10^{-2}$, $p = 0.014$). The interaction between Block Type and ROI was due to the fact that non-alternating blocks evoked more activity in the frontal regions, whereas alternating blocks evoked more activity in the temporal region ($d_{\text{temporal-frontal}} = 2.86 \times 10^{-2}$, $p = 0.007$) (Fig. 8). The three-way interaction between Block Type, ROI and Hemisphere was not significant.

The same ANOVA on the 30% compressed part of the experiment showed a main effect of ROI ($F(2, 20) = 4.391$, $p = 0.049$) and an interaction between Hemisphere and ROI ($F(2, 20) = 7.807$, $p = 0.011$), but no effect of the type of block (Fig. 9). The main effect of ROI was due to greater activation in the temporal region than in the frontal region ($d_{\text{temporal-frontal}} = 8.91 \times 10^{-3}$, $p = 0.049$, Fig. 9). This effect was modulated by Hemisphere: In the frontal region higher activity was observed in the left hemisphere, whereas the opposite pattern was observed for the temporal region (Fig. 9). The three-way interaction between Block Type, ROI and Hemisphere was not significant.

4. Discussion

In the current study, we have tested how the newborn brain perceives time-compressed speech. We compared normal speech with speech compressed to 60% and to 30% of its original duration. The former, moderately compressed speech rate is intelligible for adults, the latter, higher compression is not. We presented the stimuli using an alternating/non-alternating design and measured newborns' brain responses using NIRS.

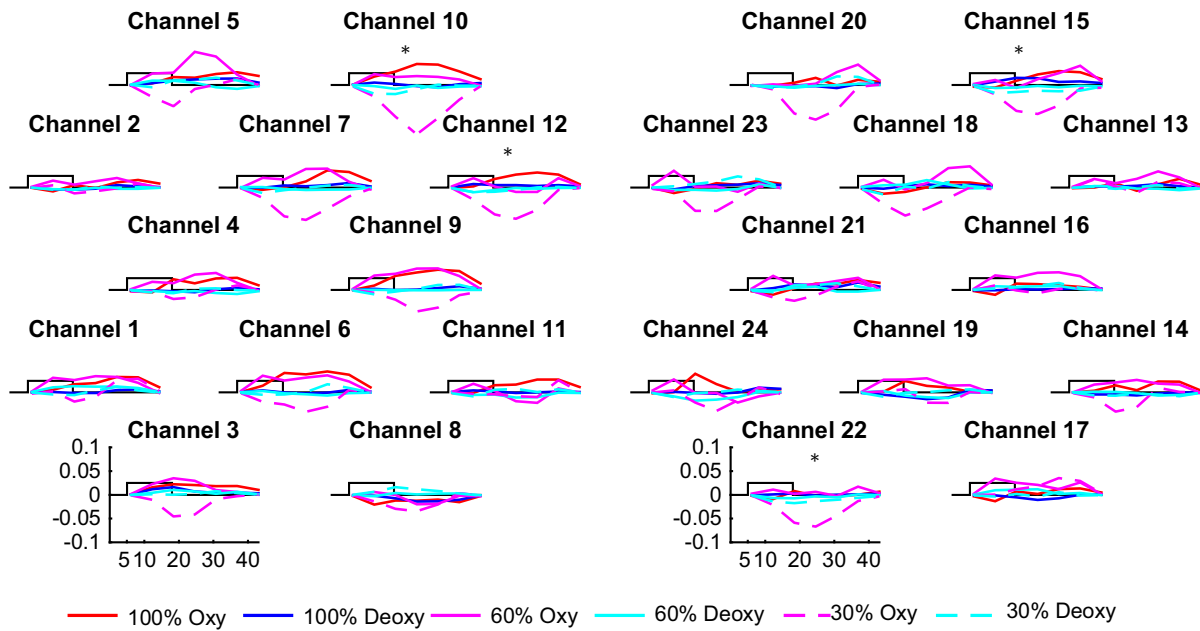


Fig. 5. Grand averages of the hemodynamic response evoked by each condition in each channel in the non-alternating blocks (mmol × mm). The rectangle represents the time of stimulation. Asterisks indicate significant comparisons for the 30% compression rate with respect to baseline.

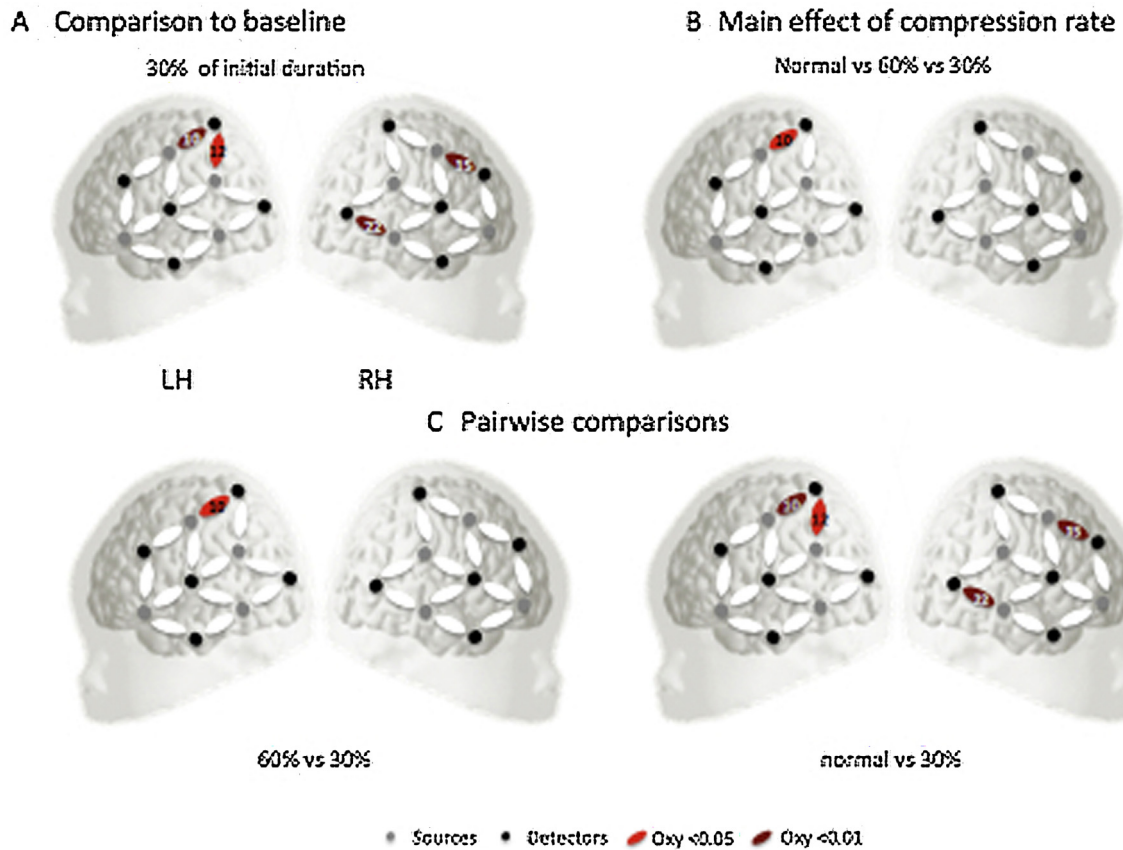


Fig. 6. Statistical maps for the non-alternating blocks. Only comparisons with at least one significant channel are shown. (A) comparison of each condition to baseline, (B) direct comparison of the three conditions, (C) pairwise comparisons of the conditions.

When comparing the neural responses to the three speech rates as presented separately in the non-alternating conditions, we have found that the fastest, unintelligible speech rate gave rise to a strong deactivation (or inverted hemodynamic response),

while the response to the two intelligible rates, i.e. the normal and the moderately compressed speech rates, was weak, but canonical. This suggests that the newborn brain processes moderately compressed speech in a similar fashion than normal speech, but

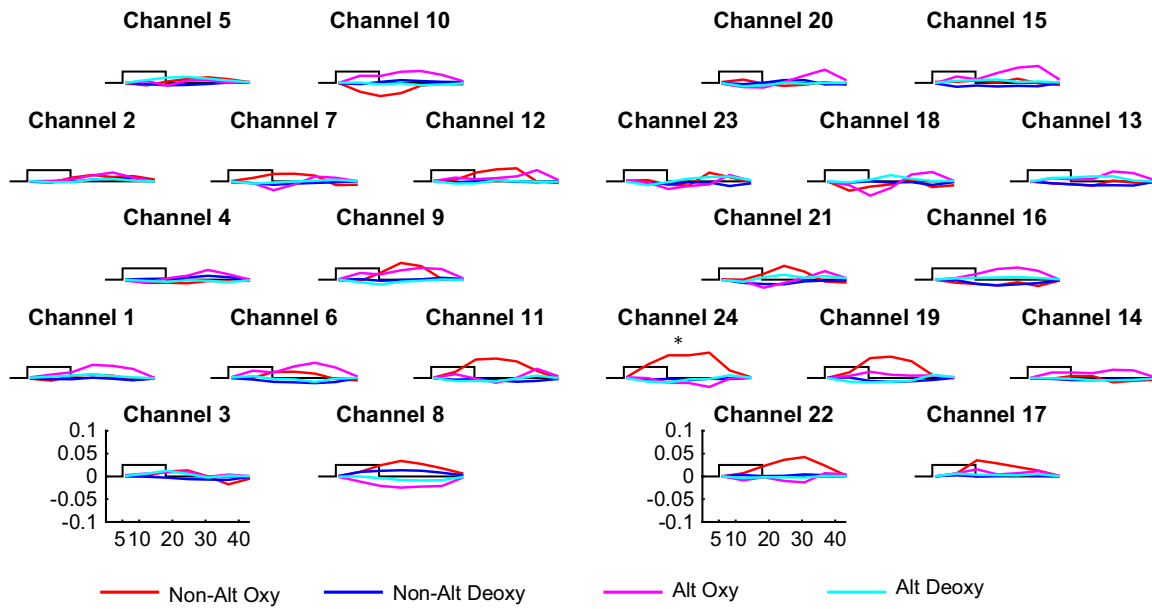


Fig. 7. Grand averages of the hemodynamic responses to the alternating and non-alternating blocks for the 60% compression rate. The rectangle represents the time of stimulation. The asterisk indicates a significant difference between alternating blocks and baseline.

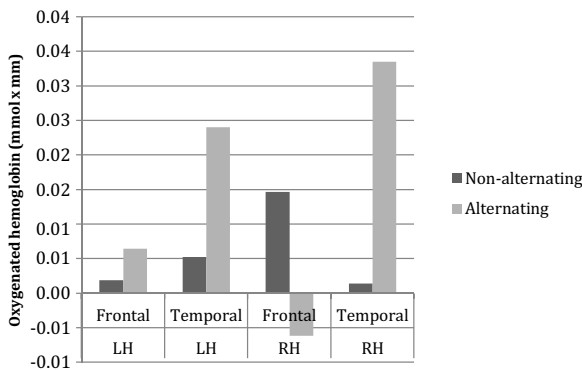


Fig. 8. Analysis by ROI for 60% compressed speech.

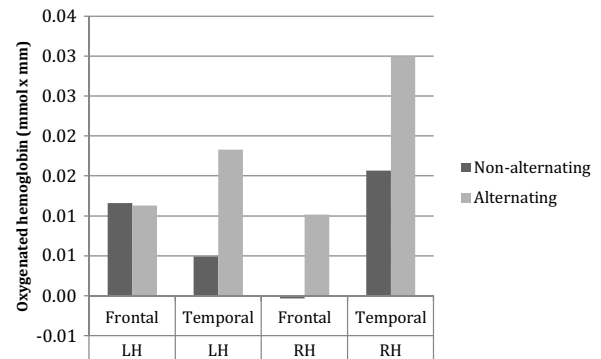


Fig. 9. Analysis by ROI for 30% compressed speech.

responds to highly compressed speech differently. This pattern of results resembles the intelligibility performance of adult listeners, who can adapt to moderately compressed, but not to highly compressed speech. Furthermore, the comparison of alternating and non-alternating blocks suggests that while the newborn brain processes normal and moderately compressed speech similarly, it is nevertheless able to discriminate between the two.

Several aspects of our results warrant further discussion before the more general implications of the findings can be considered. First, a difference between alternating and non-alternating blocks was found for the 60% compression rate, but not for the 30% compression rate alone. This can be explained when considering the pattern of responses obtained in the non-alternating blocks. Normal speech and moderately compressed speech evoked a mild canonical response, whereas the fastest speech rate evoked a strong inverted response. In the direct comparison of the alternating/non-alternating blocks, data from normal and time-compressed non-alternating blocks were averaged together before being compared to the alternating blocks. The response evoked by the normal and 30% compressed blocks averaged each other out, masking the bimodal distribution of the responses in non-alternating blocks. This made it impossible to show a difference

between the non-alternating and alternating blocks in the 30% compressed condition. By contrast, in the moderately compressed 60% speech rate, the normal and the 60% compressed non-alternating blocks evoked a similar, canonical hemodynamic response, so their mean yielded a non-null, canonical response. This pattern of results is in line with the observation that normal and moderately compressed speech rates are encoded the same way by the newborn brain, while the highest compression rate is treated differently.

We hypothesize that the relative weakness of the canonical response to the normal and the 60% compression rates is due to habituation to the stimuli across time. This habituation can be quite fast, and may happen already within the first block. While habituation across blocks is a commonly used measure in NIRS (e.g. Benavides-Varela et al., 2012; Bouchon et al., 2015), habituation within a stimulation block is hard to detect with NIRS. Existing fMRI data on adults show that adaptation to time-compressed speech occurs within the first 16 sentences (Adank and Devlin, 2010), and this number is already reached by the first alternating/non-alternating block pair. Newborns therefore might have been habituated to 60% compressed speech during the first pair of blocks, leading to reduced amplitudes between the hemodynamic responses evoked by normal and 60% compressed speech blocks. This might explain why these responses are not significantly

different from baseline. Such repetition suppression effects are often taken to be neural signatures of the existence of a stable representation of the signal (see Nordt et al., 2016, for a review). Given newborn infants' prenatal experience with their native language as well as their broad-based, universal speech perception abilities, it is not implausible to assume that they may build a stable representation of normal and moderately time-compressed speech fast and efficiently. Our experimental design was not set up to specifically test the time course of adaptation itself. A more detailed description of this process is nevertheless theoretically relevant, and will require further research.

Second, in our design, different block types contained different numbers of utterances. This choice was made to equate for the absolute amount of stimulation across block types, as this is known to influence the hemodynamic response, especially in young infants (Minagawa-Kawai et al., 2013). However, one might argue that the differences we observed across conditions might be due to this difference in the number of utterances, and not to compression rate itself. In particular, it may be the case that the inverted response observed in the fastest speech rate may be a deactivation or neural habituation effect due to the higher number of repetitions in that condition. Indeed, considerable redundancy in the stimuli has been shown to give rise to habituation effects in the NIRS response of newborns (Bouchon et al., 2015). While we did not explicitly test this alternative, there is indirect evidence in our data that this explanation is unlikely. While the non-alternating 30% compression blocks did indeed contain the greatest number of utterances (Fig. 2), alternating blocks using this compression rate also had a much higher number of utterances than non-alternating normal blocks or alternating blocks with 60% compression rate. Yet, they did not evoke an inverted response. Furthermore, the number of utterances in the different block types ranged from 6 through 7, 8 and 9 to 11. This should have resulted in graded NIRS responses for the different block types if the number of utterances played a role, which was not the case.

More generally, and returning to our research question about the origin of the mechanisms underlying adaptation to compressed speech, we had evoked several potential hypotheses: adaptation might happen at the level of (i) knowledge about the lexicon and/or grammar of the native language, (ii) processing of broadcast speech or (iii) general auditory processing requiring no experience with broadcast speech. Indeed, in the adult literature, the drop of performance at high compression ratios has been explained by several models: (i) the saturation of the lexical buffer, which gets filled up at speech rates faster than the inherent speed of linguistic read-off and processing (Vagharchakian et al., 2012), (ii) an impossibility to map phonological representations to articulatory motor plans (Adank and Devlin, 2010; Peelle et al., 2004), or (iii) a failure to find the subphonemic and rhythmic landmarks in the compressed speech signal (Pallier et al., 1998; Sebastián-Gallés et al., 2000).

Our developmental results support the third hypothesis for at least two reasons. First, despite their sophisticated abilities to process speech sounds, newborns do not have sufficient lexical or grammatical knowledge of their native language and lack experience with the processing of broadcast speech. While lexical mechanisms are certainly involved in adaptation to time-compressed speech in adults, they cannot constitute the core mechanism in newborns. Even if newborns have access to proto-lexical information, e.g. via statistical learning (Teinonen et al., 2009) or sensitivity to certain phonological features (Shi et al., 2011), they still lack a sizeable lexicon, so an explanation based on the saturation of a lexical buffer, as proposed by Vagharchakian et al. (2012), is unlikely.

Second, in adults, highly compressed and normal speech elicit different activations in the temporal cortex. Although in our study, we cannot localize the source of activation with as much precision

as in previous fMRI studies, it seems that we reproduce the posterior temporal activation found in adults (Adank and Devlin, 2010; Vagharchakian et al., 2012), as shown by the higher activation in the temporal/temporo-parietal regions for both compression rates (Figs. 8 and 9). This suggests that the neural mechanisms underlying adaptation to time-compressed speech are similar across development and are related to auditory/phonological processing.

At exactly what level of auditory processing/speech perception (from low-level acoustic processing to abstract phonological representations) this adaptation takes place remains an open question. Our study doesn't address this question directly, but on the basis of the existing literature on newborns' auditory and speech perception abilities and certain aspects of our data, we formulate possible hypotheses. First, the temporal and temporo-parietal localization of our effects suggests that prosody may play a role. These brain areas have been shown to be involved in prosodic processing in adults and in infants (Homae et al., 2006, 2007). This involvement of prosodic processing in time-compressed speech in newborns is consistent with the large body of evidence regarding newborns' extensive use of sound patterns, in particular prosody, to encode speech and language. For instance, newborns are able to recognize their native language based on its linguistic rhythm (Bertoncini et al., 1995; Nazzi et al., 1998), discriminate unknown languages if those are rhythmically different (Ramus et al., 2000) and their brain discriminates the prosodic structure of their native language from that of other languages (Abboub et al., 2016).

Indeed, in the present study babies adapted to time-compressed speech in their native language. While newborns lack experience with broadcast speech, they do have *intrauterine* experience with their native language. The intrauterine speech signal is different from the broadcast one, as maternal tissues act as low-pass filters, suppressing most individual sounds, but the low frequency modulations carrying prosody are well transmitted in the womb (Querleu et al., 1988). Even though the fetal auditory system is not fully functional yet, frequencies below 300 Hz are transmitted to the fetal inner ear (Gerhardt et al., 1992). The ability of newborns to recognize their native language (Moon et al., 1993) and its prosodic structure (Abboub et al., 2016) is evidence that they readily perceive and learn about the prosody of their native language during the end of gestation. Thus to process time-compressed speech, newborns may have relied on their knowledge of the rhythmic structure of their mother tongue to track phonological landmarks in the time-compressed signal, at least for the moderate compression rate. If newborns do indeed rely on their prenatal experience, then it is possible that they can also adapt to time-compressed speech in unfamiliar languages, provided that those belong to the same rhythmic class as their native language, just as adults can (Pallier et al., 1998). By contrast, they should not be able to adapt to time-compressed speech in a language from a different rhythmic class. Furthermore, testing adaptation to time-compression in non-linguistic sounds will make it possible to disentangle general auditory vs. speech-/language-specific mechanisms. Future research is needed to investigate these predictions.

5. Conclusions

Using moderately and highly compressed speech, we have shown that newborns are able to discriminate normal from time-compressed speech, and process normal and moderately compressed speech in similar ways. Highly compressed speech, by contrast, is treated differently. These results mirror those found in adults, bringing the first developmental evidence for the hypothesis that the ability to adapt to time-compressed speech relies on auditory, most probably prosodic/rhythmic, mechanisms.

Conflict of interest

None.

Acknowledgments

This work was supported by a Human Frontiers Science Program Young Investigator Grant nr. RGY 0073/2014, the LABEX EFL (ANR-10-LABX-0083>) and the ANR grant nr. ANR-15-CE37-0009-01 awarded to Judit Gervain. We thank the staff of the maternity of the Robert Debré Hospital for their help with participant recruitment. We also thank Lionel Granjon for his help with the stimuli, and Katie Von Holzen for her help with the permutation tests.

References

- Abboub, N., Nazzi, T., Gervain, J., 2016. Prosodic grouping at birth. *Brain Lang.* 162, 46–59 <https://doi.org/10.1016/j.bandl.2016.08.002>.
- Adank, P., Devlin, J.T., 2010. On-line plasticity in spoken sentence comprehension: adapting to time-compressed speech. *Neuroimage* 49 (1), 1124–1132 <https://doi.org/10.1016/j.neuroimage.2009.07.032>.
- Ahissar, E., Nagarajan, S., Ahissar, M., Protopapas, A., Mahncke, H., Merzenich, M.M., 2001. Speech comprehension is correlated with temporal response patterns recorded from auditory cortex. *Proc. Natl. Acad. Sci. U. S. A.* 98 (23), 13367–13372 <https://doi.org/10.1073/pnas.201400998>.
- Benavides-Varela, S., Hochmann, J.-R., Macagno, F., Nespor, M., Mehler, J., 2012. Newborn's brain activity signals the origin of word memories. *Proc. Natl. Acad. Sci. U. S. A.* 109 (44), 17908–17913 <https://doi.org/10.1073/pnas.1205413109>.
- Bertoncini, J., Floccia, C., Nazzi, T., Mehler, J., 1995. Morae and syllables: rhythmic basis of speech representations in neonates. *Lang. Speech* 38 (4), 311–329 <https://doi.org/10.1177/002383099503800401>.
- Best, C.T., Jones, C., 1998. Stimulus-alternation preference procedure to test infant speech discrimination. *Infant Behav. Dev.* 21, 295 [https://doi.org/10.1016/S0163-6383\(98\)91508-9](https://doi.org/10.1016/S0163-6383(98)91508-9).
- Bouchon, C., Nazzi, T., Gervain, J., 2015. Hemispheric asymmetries in repetition enhancement and suppression effects in the newborn brain. *PLoS One* 10 (10), e0140160 <https://doi.org/10.1371/journal.pone.0140160>.
- Dupoux, E., Green, K., 1997. Perceptual adjustment to highly compressed speech: effects of talker and rate changes. *J. Exp. Psychol. Hum. Percept. Perform.* 23 (3), 914–927 <https://doi.org/10.1037/0096-1523.23.3.914>.
- Gerhardt, K.J., Otto, R., Abrams, R.M., Colle, J.J., Burchfield, D.J., Peters, A.J.M., 1992. Cochlear microphonics recorded from fetal and newborn sheep. *Am. J. Otolaryngol.* 13 (4), 226–233 [https://doi.org/10.1016/0196-0709\(92\)90026-P](https://doi.org/10.1016/0196-0709(92)90026-P).
- Gervain, J., Macagno, F., Cogoi, S., Peña, M., Mehler, J., 2008. The neonate brain detects speech structure. *Proc. Natl. Acad. Sci.* 105 (37), 14222–14227 <https://doi.org/10.1073/pnas.0806530105>.
- 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. *Dev. Cognit. Neurosci.* 1 (1), 22–46 <https://doi.org/10.1016/j.dcn.2010.07.004>.
- Gervain, J., Berent, I., Werker, J.F., 2012. Binding at birth: the newborn brain detects identity relations and sequential position in speech. *J. Cogn. Neurosci.* 24 (3), 564–574 <https://doi.org/10.1162/jocn.a.00157>.
- Gervain, J., Werker, J.F., Geffen, M.N., 2014. Category-specific processing of scale-invariant sounds in infancy. *PLoS One* 9 (5), e96278 <https://doi.org/10.1371/journal.pone.0096278>.
- Homae, F., Watanabe, H., Nakano, T., Asakawa, K., Taga, G., 2006. The right hemisphere of sleeping infant perceives sentential prosody. *Neurosci. Res.* 54 (4), 276–280 <https://doi.org/10.1016/j.neures.2005.12.006>.
- Homae, F., Watanabe, H., Nakano, T., Taga, G., 2007. Prosodic processing in the developing brain. *Neurosci. Res.* 59 (1), 29–39 <https://doi.org/10.1016/j.neures.2007.05.005>.
- MacWhinney, B., 2000. *The CHILDES Project: Tools for Analyzing Talk, third edition*. Lawrence Erlbaum Associates, Mahwah, NJ.
- Maris, E., Oostenveld, R., 2007. Nonparametric statistical testing of EEG- and MEG-data. *J. Neurosci. Methods* 164 (1), 177–190 <https://doi.org/10.1016/j.jneumeth.2007.03.024>.
- Mehler, J., Sebastian, N., Altmann, G., Dupoux, E., Christophe, A., Pallier, C., 1993. Understanding compressed sentences: the role of rhythm and meaning. *Ann. N. Y. Acad. Sci.* 682 (1), 272–282 <https://doi.org/10.1111/j.1749-6632.1993.tb22975.x>.
- Minagawa-Kawai, Y., Cristia, A., Long, B., Vendelin, I., Hakuno, Y., Dutat, M., Dupoux, E., 2013. Insights on NIRS sensitivity from a cross-linguistic study on the emergence of phonological grammar. *Lang. Sci.* 4, 170 <https://doi.org/10.3389/fpsyg.2013.00170>.
- Moon, C., Cooper, R.P., Fifer, W.P., 1993. Two-day-olds prefer their native language. *Infant Behav. Dev.* 16 (4), 495–500 [https://doi.org/10.1016/0163-6383\(93\)80007-U](https://doi.org/10.1016/0163-6383(93)80007-U).
- Nazzi, T., Bertoncini, J., Mehler, J., 1998. Language discrimination by newborns: toward an understanding of the role of rhythm. *J. Exp. Psychol. Hum. Percept. Perform.* 24 (3), 756–766 <https://doi.org/10.1037/0096-1523.24.3.756>.
- Nichols, T.E., Holmes, A.P., 2002. *Nonparametric permutation tests for functional neuroimaging: a primer with examples*. *Hum. Brain Mapp.* 15 (1), 1–25.
- Nordt, M., Hoehl, S., Weigelt, S., 2016. The use of repetition suppression paradigms in developmental cognitive neuroscience. *Cortex* <https://doi.org/10.1016/j.cortex.2016.04.000>.
- Nourski, K.V., Reale, R.A., Oya, H., Kawasaki, H., Kovach, C.K., Chen, H., Brugge, J.F., 2009. Temporal envelope of time-compressed speech represented in the human auditory cortex. *J. Neurosci.* 29 (49), 15564–15574 <https://doi.org/10.1523/JNEUROSCI.3065-09.2009>.
- Orchik, D.J., Oelschlaeger, M.L., 1977. *Time-compressed speech discrimination in children and its relationship to articulation*. *J. Am. Audiol. Soc.* 3 (1), 37–41.
- Pallier, C., Sebastián-Gallés, N., Dupoux, E., Christophe, A., Mehler, J., 1998. Perceptual adjustment to time-compressed speech: a cross-linguistic study. *Mem. Cognit.* 26 (4), 844–851 <https://doi.org/10.3758/BF03211403>.
- Peelle, J.E., McMillan, C., Moore, P., Grossman, M., Wingfield, A., 2004. Dissociable patterns of brain activity during comprehension of rapid and syntactically complex speech: evidence from fMRI. *Brain Lang.* 91 (3), 315–325 <https://doi.org/10.1016/j.bandl.2004.05.007>.
- Querleu, D., Renard, X., Versyp, F., Paris-Delrue, L., Crèpin, G., 1988. *Fetal hearing*. *Eur. J. Obstet. Gynecol. Reprod. Biol.* 28 (3), 191–212.
- Ramus, F., Nespor, M., Mehler, J., 2000. Correlates of linguistic rhythm in the speech signal. *Cognition* 75 (1), AD3–AD30 [https://doi.org/10.1016/S0010-0277\(00\)00101-3](https://doi.org/10.1016/S0010-0277(00)00101-3).
- Rossi, S., Telkemeyer, S., Wartenburger, I., Obrig, H., 2012. Shedding light on words and sentences: near-infrared spectroscopy in language research. *Brain Lang.* 121 (2), 152–163 <https://doi.org/10.1016/j.bandl.2011.03.008>.
- Sato, Y., Sogabe, Y., Mazuka, R., 2010. Development of hemispheric specialization for lexical pitch-accent in Japanese infants. *J. Cogn. Neurosci.* 22 (11), 2503–2513.
- Sebastián-Gallés, N., Dupoux, E., Costa, A., Mehler, J., 2000. Adaptation to time-compressed speech: phonological determinants. *Percept. Psychophys.* 62 (4), 834–842 <https://doi.org/10.3758/BF03206926>.
- Shi, F., Yap, P.-T., Wu, G., Jia, H., Gilmore, J.H., Lin, W., Shen, D., 2011. Infant brain atlases from neonates to 1- and 2-year-olds. *PLoS One* 6 (4), e18746 <https://doi.org/10.1371/journal.pone.0018746>.
- Teinonen, T., Fellman, V., Näätänen, R., Alku, P., Huotilainen, M., 2009. Statistical language learning in neonates revealed by event-related brain potentials. *BMC Neurosci.* 10, 21 <https://doi.org/10.1186/1471-2202-10-21>.
- Vagharchakian, L., Dehaene-Lambertz, G., Pallier, C., Dehaene, S., 2012. A temporal bottleneck in the language comprehension network. *J. Neurosci.* 32 (26), 9089–9102 <https://doi.org/10.1523/JNEUROSCI.5685-11.2012>.