UNIVERZITA KARLOVA

Lékařská fakulta v Hradci Králové

DISERTAČNÍ PRÁCE

Josef Urbanec **2024**

Doktorský studijní program
Pediatrie

Neurofyziologické hodnocení sluchové percepce u novorozenců

Neurophysiological assessment of auditory perception in newborns

MUDr. Josef Urbanec

Školitel: doc. MUDr. Sylva Skálová, Ph.D.
Školitel konzultant: prof. Ing. Jan Kremláček, Ph.D.

Hradec Králové **2024**

Prohlášení autora

Tímto prohlašuji, že jsem tuto doktorskou disertační práci zpracoval samostatně a uvedl všechny použité informační zdroje. Zároveň dávám souhlas k tomu, aby tato práce byla uložena v Lékařské knihovně Lékařské fakulty v Hradci Králové a zde užívána ke studijním účelům za předpokladu, že každý, kdo tuto práci použije pro svou publikační nebo přednáškovou činnost, se zavazuje, že bude tento zdroj informací řádně citovat.

Souhlasím se zpřístupněním elektronické verze mé práce v informačním systému Univerzity Karlovy, Praha.

Hradec Králové, 2024 **Josef Urbanec**

Poděkování

Velmi rád bych tímto poděkoval své školitelce paní docentce Skálové za projevenou důvěru a umožnění mého postgraduálního studia. Velké vyjádření díků patří mému školiteli konzultantovi panu profesorovi Kremláčkovi a kolegyni doktorce Chládkové. Jejich odborný, a především lidský přístup mi několikrát pomohl najít v sobě novou sílu na dokončení tohoto studia. V neposlední řadě bych rád poděkoval zdravotním sestřičkám na gynekologicko-porodnickém oddělení v Hradci Králové a Havlíčkově Brodě za jejich obětavost, dobrou náladu a pomoc při náboru dětí do novorozeneckých studií.

Grant

Tento výzkum byl podpořen projekty Univerzity Karlovy PRIMUS/17/HUM/19 a PROGRES Q40/07.

Abstrakt

Cílem naší práce bylo posouzení sluchové percepce u donošených i předčasně narozených novorozenců měřením korových evokovaných potenciálů. Povrchovými elektrodami umístěnými na hlavě novorozence byl během zvukové stimulace snímán elektroencefalografický (EEG) záznam. Kortikální aktivita a její vztah ke sluchové percepci byly studovány pomocí tzv. na signál vázaných evokovaných potenciálů (event-related potencials, ERP). Vzhledem k nevyzrálosti centrálního nervového systému (CNS) dětí byly dále vyhodnocovány a posuzovány odvozené komponenty především mismatch-response (MMR), které jsou zcela nezávislé na pozornosti zkoumaného subjektu a vybavitelné ve spánku.

Před plánovanými novorozeneckými studiemi proběhl pilotní výzkumný projekt se zdravými dospělými dobrovolníky. Během této části projektu byla ověřena připravená zvuková stimulace. Hláskové a tónové kontrasty zvukové stimulace vyvolaly u dospělých jedinců dostatečné kortikální odpovědi objektivizované pomocí ERP. Výsledky této studie jsou uvedeny níže v kapitole 4.1. Ověřená zvuková stimulace byla následně použita při výzkumu sluchové percepce novorozenců.

Hlavní částí předkládaného výzkumu byla elektrofyziologická studie sluchové percepce se zaměřením na diskriminaci hláskových a tónových kontrastů u zralých novorozenců v prvních dnech po narození. Pomocí předem vybraných 6 skalpových elektrod (F3, F4, C3, C4, Fz, Cz), snímajících kortikální odpověď na přicházející zvukový signál, byl pořizován EEG záznam. Zprůměrováním tohoto záznamu jsme pro jednotlivé elektrody získaly křivky ERP, respektive mismatch-response (MMR). Zjistili jsme, že novorozenci jsou schopni lépe rozlišovat hláskové kontrasty v porovnání s tónovými kontrasty stejných fyzikálních vlastností. MMR byla lépe vyjádřená (tedy s negativní amplitudou) a levostranně detekovatelná pro změnu délky trvání hlásek [fe] -> [fe:] i změnu spektrální kvality hlásek [fa] -> [fe]. Toto robustněji vybavitelné vnímání hláskových kontrastů vykazovalo zralost CNS podobné dospělému člověku ve zpracování právě řečových podnětů, relativně k nezralému zpracování neřečových podnětů – tónů.

Na výše uvedená zjištění navázala studie porovnávající sluchovou percepci u lehce nezralých a donošených novorozenců. Tato studie se již výhradně zaměřovala na vnímání hlásek, tónové nahrávky nebyly použity pro nedostatečnou odpověď u zralých novorozenců v předešlé studii. Výsledky potvrdily, že vyzrávání sluchové percepce je přímo úměrné rostoucímu gestačnímu věku. Předčasně narození novorozenci měli sníženou odpověď na zvukovou stimulaci v porovnání s donošenými novorozenci. Hranice vyzrálé, dospělým jedincům podobné sluchové percepce se tak shoduje s všeobecně uznávanou hranicí gestační zralosti, tj. 37. týdnem těhotenství.

Klíčová slova: kognitivní sluchové potenciály, evokované potenciály, mismatch-negativity, sluchová percepce novorozenců

Abstract

The aim of our research was to assess auditory perception in term neonates by measuring cortical evoked potentials. An electroencephalographic (EEG) recording was made during sound stimulation through surface electrodes placed on the newborn's head. Cortical activity and its relationship to auditory perception were studied using so-called event-related evoked potentials (ERPs). Due to the immaturity of the central nervous system (CNS) of infants, derived modalities, mainly the mismatch-response (MMR), which is completely independent of the attention of the subject, were further evaluated and assessed.

Prior to the planned neonatal study, a pilot study was conducted with healthy adult volunteers. During this part of the project, the prepared sound stimulation was verified. Vowel and tonal contrasts of the sound stimulation elicited sufficient cortical responses objectified by ERPs in adults. The results of this study are presented in section 4.1 below. The verified sound stimulation was then used in studies with newborns.

The main part of the presented research was an electrophysiological study of auditory perception with a focus on the discrimination of speech and tonal contrasts in mature newborns in first days after birth. An EEG recording was made using 6 scalp electrodes (F3, F4, C3, C4, Fz, Cz) to measure the cortical response to the incoming sound signal. By averaging this recording, we obtained ERP and MMR curves for individual electrodes. We found that newborns can better discriminate phonetic contrasts than tonal contrasts of the same physical properties. Furthermore, the MMR was larger (i.e. with a negative amplitude) and detectable on the left side for the change in vowel duration [fe] -> [fe:] and the change in vowel spectrum [fa] -> [fe]. This robust perception of speech contrasts, compared to the perception of non-speech sounds, indicated a relatively mature CNS for the perception of speech sounds specifically.

The above results were closely followed by a study comparing auditory perception in mild preterm and term newborns. This study already focused exclusively on the perception of voices. Tonal recordings were not used after the previous study had found an insufficient response in mature newborns. The results confirmed that the maturation of auditory perception is directly proportional to increasing weeks of gestation. Premature infants had a reduced response to sound stimulation compared to full-term infants. The limit of mature, adult-like auditory perception thus coincides with the generally accepted limit of gestational maturity, i.e. the 37th week of gestation.

Key words: Cognitive Auditory Potentials, Evoked Potentials, Mismatch-negativity, Auditory Perception of Newborns

OBSAH

1. SEZNAM POUŽITÝCH ZKRATEK

AABR	automatic auditory brainstem responses
APD	auditory processing deficit
AUC	area under curve
CAEP	cortical auditory evoked potentials
CNS	centrální nervová soustava
dB	decibel
EEG	elektroencefalografie
ERPs	event-related potencials
MMN	mismatch-negativity
MMR	mismatch-response
TEOAE	transient evoked otoacustic emissions
C3	centrální, párová elektroda umístěna vlevo od střední čáry
C4	centrální, párová elektroda umístěna vpravo od střední čáry
Cz	centrální, nepárová elektroda umístěna ve střední čáře nad vertexem
F3	frontální, párová elektroda umístěna vlevo od střední čáry
F4	frontální, párová elektroda umístěna vpravo od střední čáry
Fz	frontální, nepárová elektroda umístěna ve střední čáře nad vertexem
Hz	hertz
ms	milisekunda
P300	pozitivní evokovaný potenciál detekovatelný s latencí 300 ms od stimulace
N400	negativní evokovaný potenciál detekovatelný s latencí 400 ms od stimulace

2. ÚVOD

Sluch je jedním ze základních smyslů potřebných pro správný vývoj řeči stejně jako ostatních kognitivních funkcí např. vývoj abstraktního myšlení. Adekvátní sluchová percepce je nutná k optimálnímu psychomotorickému vývoji každého dítěte[1,2]. Novorozenecké období plynule navazuje na nitroděložní vývoj. Během vývoje je plod od konce 2. trimestru stimulován zvukovými podněty, tedy mechanickým vlněním procházejícím děložní stěnou a plodovými obaly[3]. I značně modifikované zvuky stimulují sluchovou dráhu plodu[4]. Šíří se v těle plodu především kostním vedením, zatímco po narození převáží jeho vzdušné vedení. Pro jejich zpracování je nezbytným předpokladem správně vyvinutá sluchová dráha[5-8]. Proto bylo vypracováno několik vyšetřovacích metod, které dokáží po narození dítěte zhodnotit funkčnost sluchové dráhy. Mezi ty nejrozšířenější patří vyšetření otoakustických emisí (transient evoked otoacustic emissions, TEOAE) nebo krátkodobých evokovaných potenciálů mozkového kmene (automatic auditory brainstem responses, AABR)[5,9]. Tyto metody dokáží zhodnotit funkčnost pouze části převodního systému sluchové dráhy a její nižší, podkorová centra. Abychom mohli podrobněji vyhodnotit percepci zvuku a jeho zpracování vyššími, korovými centry, využíváme dlouhodobých evokovaných potenciálů. Jejich vyšetření je mnohem citlivější a odráží nejen šíření zvukového signálu sluchovou dráhou, ale také jeho zpracování primární i asociovanými sluchovými korovými oblastmi. Podrobnější vyšetření může odhalit funkční poruchy sluchové kognice (auditory processing deficit, APD), mezi které řadíme dyslexii či vývojové poruchy učení[10,11]. K vyšetření těchto center využíváme metodu dlouhodobých neboli kognitivních, na událost vázaných potenciálů, které hodnotí funkční integritu sluchové dráhy jako reakci mozkové kůry na definované zvukové podněty[12,13].

2.1 KOGNITIVNÍ ERPs

Sluchové podněty (stejně tak např. i zrakové) vyvolávají neurální elektrickou aktivitu CNS. Tuto aktivitu můžeme objektivně změřit dynamickou metodou EEG během expozice definovanými akustickými podněty[13,14]. Získaná data jsou následně zpracována a vyhodnocena jako evokované potenciály. Odpověď mozkové kůry je detektovatelná s nejdelší latencí po zvukové stimulaci v řádu stovek milisekund po zaznění stimulu. Jedná se tak o dlouhodobé evokované potenciály. Kortikální sluchové potenciály (cortical auditory evoked potentials, CAEP) patří mezi endogenní sluchové potenciály. Endogenní potenciály nejsou ovlivněny pouhými fyzikálními vlastnosti zvuku (zde by se jednalo o exogenní potenciály), ale předpokládá se jejich zpracování a ovlivnění vyššími korovými centry. Endogenní komponenty jsou vázány na očekávaný zvukový podnět a jeho komplexní zpracování mozkovou kůrou. Je možné je využít k hodnocení schopnosti rozlišit komplexní charakteristiky zvukových stop. Mezi nejčastěji využívané endogenní komponenty evokovaných potenciálů řadíme vlnu P300, N400 a odvozenou MMR[15]. Tyto komponenty hodnotíme na základě časového rozlišení pomocí vrcholového času (latence), jeho polarity a amplitudy[12]. MMR je určena rozdílem amplitudy napětí potenciálu vyvolaného deviantním (vzácně se vyskytujícím) zvukovým podnětem a potenciálu standardního (často se opakujícího) podnětu. Deviantní zvuková stimulace vyvolá chybnou aktivitu korové oblasti původně predikované na základě předcházející zkušenosti s dlouhodobě opakovaným vjemem. Rozdíl těchto odpovědí je zachycen s latencí 100-250 ms po výskytu deviantního stimulu[14]. MMR má u dospělých jedinců obvykle negativní polaritu a můžeme ji nazvat mismatch-negativity (MMN). Polarita MMR se mění s věkem, její postupná změna z pozitivní na negativní polaritu je odrazem dozrávání CNS u dětí a časných dospělých. Vývojově je MMR velmi raný potenciál detekovatelný již od 30. gestačního týdne umožňující intrauterinně zpravovat přicházející zvukové signály[13,16]. Vzhledem k této skutečnosti je MMR vhodná k posouzení neurální aktivity CNS, tedy i sluchové percepce novorozence[5,10]. Její měření nevyžaduje aktivní pozornost a je ideální ho provádět během spánku dítěte k eliminaci pohybových i jiných artefaktů.

Komplexní souhrn výše zmíněné problematiky a základní charakteristika jsou shrnuty v přehledovém článku „The development and neurophysiological assessment of newborn auditory cognition: a review of findings and their application“ publikovaném v časopise Acta medica (Hradec Králové) viz následující kapitola 2.2.

2.2 The Development and Neurophysiological Assessment of Newborn Auditory Cognition: A Review of Findings and Their Application

Josef Urbanec[1,2,*], Jan Kremláček[1,3], Kateřina Chládková[4,5], Sylva Skálová[6]

AUTHOR AFFILIATIONS

1 Department of Pathological Physiology, Medical Faculty in Hradec Králové, Charles University, Czech Republic

2 Paediatrics Department, Havlíčkův Brod Hospital, Czech Republic

3 Department of Medical Biophysics, Medical Faculty in Hradec Králové, Charles University, Czech Republic

4 Institute of Psychology, Czech Academy of Sciences, Prague, Czech Republic

5 Institute of Czech Language and Theory of Communication, Faculty of Arts, Charles University, Prague, Czech Republic

6 Paediatrics Department of University Hospital in Hradec Králové, Charles University, Czech Republic

* Corresponding author: Department of Pathological Physiology, Medical Faculty in Hradec Králové, Charles University, Czech Republic; e-mail: jurbanec86@gmail.com

ABSTRACT

This review article introduces the basic principles of infants' neurophysiology, while summarizing the core knowledge of the anatomical structure of the auditory pathway, and presents previous findings on newborns' neural speech processing and suggests their possible applications for clinical practice. In order to tap into the functioning of the auditory pathway in newborns, recent approaches have employed electrophysiological techniques that measure electrical activity of the brain. The neural processing of an incoming auditory stimulus is objectively reflected by means of auditory event-related potentials. The newborn's nervous system processes the incoming sound, and the associated electrical activity of the brain is measured and extracted as components characterized by amplitude, latency, and polarity. Based on the parameters of event-related potentials, it is possible to assess the maturity of a child's brain, or to identify a pathology that needs to be treated or mitigated. For instance, in children with a cochlear implant, auditory event-related potentials are employed to evaluate an outcome of the implantation procedure and to monitor the development of hearing. Event-related potentials turn out to be an irreplaceable part of neurodevelopmental care for high-risk children e.g., preterm babies, children with learning disabilities, autism and many other risk factors.

KEYWORDS

newborns; auditory pathway; cortical auditory evoked potentials; maturation of the central nervous system; learning disabilities

INTRODUCTION

The neonatal period is defined as the interval from birth to the 28th day of an infant's life. Despite being marked by its beginning and end points, the neonatal period should – in many respects – be understood as a direct continuation of intrauterine development. According to knowledge of auditory perception, it is well-established that the fetus can hear and process surrounding stimuli and adequate prenatal auditory stimulation is necessary for normal development of hearing (1, 2).

After birth, hearing becomes one of the fundamental senses that stimulate the early development of a child's cognitive functions, thus contributing to the acquisition of speech, language, and abstract thinking. Intact peripheral and central part of the auditory apparatus is necessary for a child's psychomotor development. As hearing impairment may interfere with cognitive and psychomotor development, it is crucial to detect this deficit as soon as possible. Subsequent intervention, e.g. with a cochlear implant (CI), may reduce impact on all aspects of later life quality (3–7). For this reason, objective screening methods focused on auditory perception are typically performed. The most common is the assessment of transient evoked otoacoustic emissions (TEOAE). This approach can assess the functionality of cochlea (the peripheral part of the auditory apparatus) but cannot measure whether the information has also been correctly processed by the central nervous system (CNS). Improper engagement and functioning of the higher auditory areas can lead to disorders such as the auditory processing deficit, dyslexia, or learning disability (3, 8). Detection of the brainstem, early, and later evoked potentials, also called event-related potentials (ERPs), allow us to examine the subsequent stages of auditory stimulus processing. These techniques objectively test the functional integrity of the auditory system by measuring the brain's response to auditory stimuli (9).

ANATOMY OF AUDITORY PATHWAY

The auditory pathway is distinguished into the peripheral and the central part, also called structural and neurosensorial, respectively (Figure 1). These two parts differ not only in their function, but also in the timeline of their development. The peripheral part consists of the outer, middle, and inner ear. It participates in capturing and converting an incoming auditory stimulus (mechanical sound waves) into electrical potential, which is transferred to the central auditory system (1). The division of the peripheral system into the outer, middle, and inner ear mostly follows the development of primary germ layers or their derivatives (Figure 2A–D). The base of the inner ear forms at the beginning of the fourth gestational week and its development completes in the 20th gestational week (1, 10, 11).

It is through the vestibulocochlear nerve that the auditory receptor potential reaches the brainstem, afterwards switching to the mesencephalon, thalamus, and finally the cerebral cortex. The primary auditory cortex is in the temporal lobe, in the tonotopically arranged area 41 (Figure 1). The axons end in the associative cortical regions areas 42 and 22. This part of the auditory system does not develop fully until the 20th gestational week (12, 13).

The cochlea of the inner ear and the auditory cortical networks in the temporal lobe are, developmentally, the most sensitive clinical components of the auditory pathway. They may be affected during intrauterine development, e.g. by prenatal infection, but also in the neonatal period due to antibiotic treatment, or exposure to noise in a neonatal intensive care unit (14). This vulnerability stems largely from the gradual maturation of the sensitive neurosensory part (the hair cells of the inner ear), axons and neurons, that takes place between the 25th gestational week and the fifth month of life (1).

The auditory pathway can transmit the surrounding sound stimuli to the developing fetal brain already between the 25th and the 29th gestational week. During gestation, the uterus is a natural barrier protecting the fetus from intensive impacts that could harm its development, limiting the intensity as well as the spectral content of the incoming sound (1, 3, 15). However, even in the rather attenuated and somewhat distorted sound, a physiologically developing fetus can recognize various frequently encountered sounds, most notably the rhythm and melody of its mother's speech (16). Prenatal auditory stimulation aids the development of the tonotopic organization of the cochlear hair cells and the auditory cortex (14). After birth, when the attenuating barrier disappears, the incoming auditory stimuli contribute to further cortical development. From the perspective of hearing, the neonatal period is an uninterrupted continuation of intrauterine development (1, 2). This is evidenced by a study that compared the development of hearing with vision. While vision develops only after birth, auditory stimulation with varied naturalistic stimuli (e.g. maternal voice, music, or common environmental sounds) during the last 10–12 weeks of the fetal period in utero or in prematurely born infants seems to be essential for proper hearing development (1).

CORTICAL EVOKED POTENTIALS

Neuronal activity induced by auditory stimulation can be detected as evoked potentials, at many different levels of the auditory pathway. The measurement of evoked potentials is a non-invasive, dynamic, and objective method based on the principle of electroencephalography (EEG) sensing the electrical activity of the brain. Cortical Auditory Evoked Potentials (CAEPs) are often measured to assess auditory perception. They belong to a broader group of ERPs, sometimes called cognitive ERPs (9). ERPs extraction is done by averaging epochs of the EEG that are aligned to the occurrence of repeatedly presented acoustic stimuli (12, 17).

To assess the trajectory of auditory processing one typically evaluates the components, i.e. the peaks and their latencies, within the averaged ERPs. The advantage of the ERP method is its fine temporal resolution, which allows to accurately measure the peak time of a response, i.e., the latency, in milliseconds (9). The strongest CAEPs can be recorded in the back of lateral sulcus, the so-called Sylvian fissure, which separates the frontal and temporal lobes. Due to the non-invasive character of EEG recording the exact localization of CAEPs is not possible (12, 17).

With some simplification, CAEPs can be divided into exogenous (sometimes inaccurately called obligatory) and endogenous (inaccurately called cognitive) components. Exogenous components reflect the physical properties of the sound, such as the intensity, frequency, and duration, whereas endogenous components are modulated by neuronal activity in higher cortical centres and are not determined solely by the sound's physical properties (17).

Exogenous components include the P50, N100, P200, and N200. In newborns, unlike in older children, P100 and N100 waves are not well detectable. Newborns' ERPs typically have a relatively broad peak at 200–300 ms latency, called P200, which is followed by a broad negative N200 wave at 300–600 ms latency. The latencies and breadth of the P200 and N200 waves decrease markedly in the course of the first months after birth (9, 12). Endogenous components are used to evaluate higher-level, e.g. linguistic, processing of auditory stimuli by the newborn brain. These components include the mismatch response (MMR) (18), P300, and

N400. MMR, one of the most frequently evaluated components, is defined as a difference in the potential induced by a rarely occurring, i.e. deviant, stimulus, and the potential induced by a frequently repeated, i.e. standard, stimulus (Figure 3). The MMR is roughly interpretable as an index of prediction error originating from a comparison of a novel unexpected deviant stimulus against a built-up memory trace for the previously presented frequent standard stimuli (12). The MMR component is elicited automatically and does not require conscious attention to the stimuli, and can be also measured during (active) sleep. If a deviant sound is perceived as different from previously presented standard sounds, it elicits the MMR, typically at a latency of 100–250 ms relative to the onset of the deviation. The larger the perceived difference between the deviant and the standard stimulus, the larger the MMR amplitude and/ or the shorter its latency. In adults, the MMR is typically bilateral in both temporal and frontal cortical areas (12) and has a negative polarity (hence in adults it is referred to as mismatch negativity, MMN, see Figure 3). In infants, however, MMR often has a positive polarity (3), indicating imperfect maturation and/or marginal audibility of the acoustic difference between the deviant and the standard stimulus (4).

Besides the age-related differential polarity, the MMR latency is in newborns greater than in adults and decreases gradually mainly during the first two years of life. Ontogenetically, the MMR is a very early potential detectable from the 30th postconceptional week (14, 17). Newborns' MMR, similarly to adults' MMN, reflects rather fine phonetic discrimination abilities, such as the ability to distinguish sounds coming from different sources, or the ability to detect both a change in speaker voice and in speech sound quality (9). This observation in healthy newborns indicates that the neonatal brain has a fully developed discriminatory capacity for sound stimuli (17), although its CNS structures are not yet fully mature (19–21). Newborns' MMR also indexes the ability to differentiate variations in auditory stimuli that are important for speech and language development (17). In child auditory perception, developmental speech disorders or learning difficulties are often associated with an attenuated or delayed MMR response (3). MMR is therefore well suited to assess the earliest stages of cognitive development, particularly the speech and language capacity of the developing individual.

STUDIES WITH NEWBORNS

Several studies have assessed and evaluated auditory cognitive potentials in neonates. Most of studies test healthy newborns and apply inclusion criteria such as the absence of neurological disorders, medication, pre- or peripartal complications, excessive physical activity during the assessment, and need a passed neonatal hearing screening – brainstem auditory evoked potentials, steady state response auditors or TEOAE (4, 19). In previous studies, healthy newborns meeting the above criteria are typically compared to e.g. preterm newborns, infants with suspicion of hearing impairment, deficient neural speech processing, or high familial risk for a developmental language or speech disorder.

Melo et al. (2016) compared the cognitive evoked potentials of 31 preterm and 66 term infants. The infants were tested in sleep, after feeding, using biaural auditory stimulation. The syllable /ba/ served as the frequent standard stimulus, and /ra/ served as the rare deviant stimulus. The P100 and N100 waves were less likely to be present in preterm as compared to full term infants (they were missing in 13% and 4.5% of cases, respectively). No Fig. 3 Schematic representation of cortical auditory evoked potentials (CAEP) sensed by an electrode placed above the frontal area (Fz) and the processus mastoideus (M2). The frequent, standard stimulus is represented by a green curve, the rare, deviant stimulus by an orange curve. The subsequent amplitude difference of both stimuli is highlighted by a blue curve as the so-called difference wave, which peaks as mismatch negativity (MMN) at latency of about 200 ms. The amplitude of the MMN tends to be positive when measured with an electrode above the mastoid processus, in other locations it typically, in adults, has negative values (scheme adopted and freely adjusted according to (17)). significant differences in the incidence of N200 or P200 were found between the two groups. The absence of the P100 wave in CAEP in premature infants can be a possible indicator of cognitive delays or immature cortical structures in this population. Besides evaluating the absence/ presence of P100 (and N100), the latency of ERPs components can, be used too as an indicator of immaturity inversely proportional to gestational age (4).

The results of that study are in line with the results of other studies comparing the maturation of the infant brain. Exogenous components have longer latency in newborns than in older children, and the latency rapidly decreases in the first and second year of life. This may be caused by the development of synapses during the first years of life, reflected in an increase of low-frequency EEG activity, which is also the frequency range relevant for the ERPs. Continuing myelination at pre-school age leads to more adult-like ERPs.

In general, ERP latency thus mostly reflects the maturation of the CNS itself. ERP amplitude, on the contrary, seems to correlate with the number of neural structures involved in the response (number of synapses). Early developmental changes in the amplitude of the auditory ERP thus seem to depend mainly on gestational age, and less so on the amount of (extrauterine) auditory exposure (2, 4, 20, 21).

A recent study by Oliveira et al. (2019) assessed CAEPs in 39 full-term newborns (19). The measurements were monoaural with a randomly selected ear stimulated by pure tones of various frequencies. At an initial sound intensity of 80 dB SPL, latency and amplitude did not show statistically significant differences for various stimulus frequencies. However, the latency of the P100 wave was inversely proportional to stimulus intensity. One of the conclusions of this study was that compared to the brain stem response, the cortical auditory ERPs are elicited only if stimulus intensity exceeds a particular threshold (2, 19). The fact that the brain stem response is elicited also at a lower stimulus intensity can be attributed to a faster maturation of the subcortical, compared to cortical centres. Some other studies found that the latencies of P100 and N100 are greater for pure tones than for speech stimuli (19, 22).

ERPs can be used not only to assess CNS maturation, but also to quantify the success of intervention in children with hearing disorders, especially with deafness. Silva et al. (2014) have shown that auditory cognitive potentials can verify the level of auditory stimulation needed for the maturation of the CNS in children with CI. For instance, there seems to be a relationship between the P100 wave, measured immediately after CI implantation, and the onset of vocalisation in children with different ages of CI implantation (6). After implantation, which positively affects the child's communicative development, one can objectively assess changes in the CNS, namely, a decrease of the P100 latency to tones and speech stimuli (4–7).

The CAEPs may assess the effect of CI implantation and normalization of auditory development but could also detect deafness in children. Mehta et al. (2017) described the role of the CAEPs for early diagnosis and later therapy in children with hearing loss in United Kingdom during 2011–2015. That study compared 2 sequential cohorts of children with a permanent childhood hearing impairment and with different time of CI implantation. The first cohort included 34 children examined prior the introduction of CAEPs, the second 44 children examined after the introduction of CAEPs. The only difference in the patient pathway was the use of CAEPs in diagnosis and therapy. Except the common examination, for the second infants group diagnosis included CAEPs to speech tokens /m/ (duration of 30 ms), /g/ (duration 20 ms), and /t/ (duration of 30 ms) presented at nominal intensity 55, 65 and 75 dB SPL. Early hearing aid fitting was recommended if the response for /g/ or /t/ at 55 dB SPL was missing. Additionally, a second CAEPs session 4 to 8 weeks later was performed for all children without a recommendation of early hearing aid at the first session. If the CAEPs (at second session) were absent at 75 dB SPL in infants optimally fitted with hearing aids, referral for CI assessment was recommended. The results showed that children with severe deafness were referred significantly earlier for CI assessment after the introduction of CAEPs than before: the median age of hearing aid fitting for children with all degrees of hearing impairment decreased from 9.2 months to 3.9 months after the introduction of CAEPs examination. This trend was observed also in children with mild or moderate hearing loss (median age decreased from 19 to 5 months) (7).

There are other areas in which CAEPs seem promising as an early diagnostic tool for developmental disorders. Thiede et al. (2019) performed a longitudinal study with 44 newborns at high familial risk of dyslexia and with a control group of 44 low-risk newborns. The newborns were stimulated by pseudowords with changes from a standard /tata/ stimulus in vowel duration /tata:/, vowel spectrum /tato/ and pitch /tata/ at stimulus intensity 65 dB SPL. EEG recordings were analysed for MMR to each type of change. The results suggested atypical neural discrimination of speech sound differences in the high-risk newborns: their MMR were diminished or completely absent, had longer latency and different hemispheric lateralization and morphology compared to infants with no dyslexia in family history (3).

CONCLUSIONS AND CLINICAL APPLICATION

The auditory pathway is a necessary and irreplaceable connection of the developing fetus with the outside world. The peripheral and central auditory system development starts already in the prenatal period and at birth, hearing seems comparable in pre-term and term neonates (4). At the 40th gestational week, auditory cognitive potentials of premature and term-born infants do not seem to differ significantly, indicating that extrauterine stimulation does not alter the maturation of auditory processes in the pre- and postnatal period (17). Auditory ERPs display maturational changes throughout infants' development. Throughout infancy there is a clear developmental decrease in latency which is comparable across children born premature and children born full-term (same gestational age), despite the former group having had longer exposure to sounds ex utero, which aligns well with the gradual maturation of CNS structures across the intrauterine and extrauterine periods of development (19, 21).

The absence or reduced amplitude of ERP components can be used for diagnosis and evaluation of pathologies. As an example, MMR deficiency is often associated with learning disorders, cleft palate, autism or Asperger syndrome, depression or behavioural disorders. In children with very low birth weight and speech impairment, reduced MMR amplitude was found at four to six years of age (9). This reduction in MMR amplitude is to be associated with speech impairment rather than with the child's maturation at birth because, as noted above, the amplitude and latency of the measured cognitive potential components are comparable between term and very-low-birthweight (premature) children (4, 17).

To conclude, electrophysiological methods are routinely employed to monitor neonatal hearing but here we show that they could have a greater application in the clinical practice as they can help assess the very development and maturation of the newborns' auditory pathway.

Maturation of CNS depends primarily on the myelination of nerve fibers, which lead the signal to the corresponding Newborn Auditory Cognition: A Review 7 cortical centres which generate the cortical evoked potentials (19). Moreover, early and developmental evaluation of auditory ERPs is a promising approach that may find application in monitoring the dynamics of some developmental disorders and diseases such as dyslexia, autism (3, 8, 14). Based on recent findings which were reviewed in this article, we suggest that CAEPs should become an integral part of clinical practice to evaluate children's auditory development.

DEDICATION

Supported by the projects of Charles University PRIMUS/17/HUM/19 a PROGRES Q40/07.

REFERENCES

1. Graven S, Browne J. Auditory Development in the Fetus and Infant. NbInfant Nurs Rev 2008; 8(4): 187–93.

2. Sousa A, Didoné D, Sleifer P. Longitudinal Comparison of Auditory Steady-State Evoked Potentials in Preterm and Term Infants: The Maturation Process. Int Arch Otorhinolaryngol 2017; 21(3): 200–5.

3. Thiede A, Virtala P, Ala-Kurikka I, et al. An extensive pattern of atypical neural speech-sound discrimination in newborns at risk of dyslexia. Clin Neurophysiol 2019; 130(5): 634–46.

4. Melo de A, Biaggio E, Rechia I, et al. Cortical auditory evoked potentials in full-term and preterm neonates. Codas 2016; 28(5): 491–6.

5. Martins K, Gil D. Cortical Auditory Evoked Potentials with Simple (Tone Burst) and Complex (Speech) Stimuli in Children with Cochlear Implant. Int Arc Otorhinolaryngol 2017; 21(4): 351–7.

6. Silva L, Couto M, Tsuji R, et al. Auditory pathways' maturation after cochlear implant via cortical auditory evoked potentials. Braz J Otorhinolaryngol 2014; 80(2): 131–7.

7. Mehta K, Watkin P, BAldwin M, et al. Role of Cortical Auditory Evoked Potentials in Reducing the Age at Hearing Aid Fitting in Children With Hearing Loss Identified by Newborn Hearing Screening. Trends in Hearing 2017; 21.

8. Frizzo, A. Auditory evoked potential: a proposal for further evaluation in children with learning disabilities. Front Psychol 2015; 6: 788.

9. Duncan C, Barry R, Connolly J, et al. Event-related potentials in clinical research: Guidelines for eliciting, recording, and quantifying mismatch negativity, P300, and N400. Clin Neurophysiol 2009; 120(11): 1883–908.

10. Vacek Z. Organogeneze. Embryologie: učebnice pro studenty lékařství a oborů všeobecná sestra a porodní asistentka. Praha: Grada 2006: 99–101, 235–8.

11. Carlson, M. Nervous system. In: Human embryology and developmental biology, 5th ed. Philadelphia, PA: Elsevier/Saunders 2014: 216–45.

12. Joos K, Gilles A, Van de Heyning P, et al. From sensation to percept: The neural signature of auditory event-related potentials. Neurosci Biobehav R 2014; 42: 148–56.

13. Druga R, Grim M, Dubový P. Přehled drah CNS. Anatomie CNS. Praha: Galén 2011: 208–10.

14. Lahav A, Skoe E. An acoustic gap between the NICU and womb: a potential risk for compromised neuroplasticity of the auditory system in preterm infants. Front Neurosci 2014; 8: 381.

15. Wodicka G, M. Lam A, Bhargava V, et al. Acoustic impedance of the maternal abdomen. J Acoust Soc Am 1993; 94(1): 13–18.

16. Granier-Deferre C, Ribeiro A, Jacquet AY, et al. Near-term fetuses process temporal features of speech. Dev Sci 2011; 14(2): 336–52.

17. Fellman V, Huotilainen M. Cortical auditory event-related potentials in newborn infants. Semin Fetal Neonatal Med 2006; 11(6): 452–8.

18. Naatanen R, Picton T. The N1 wave of the human electric and magnetic response to sound: a review and an analysis of the component structure. Psychophysiology 1987; 24(4): 375–425.

19. Oliveira L, Didoné D, Durante A. Automated cortical auditory evoked potentials threshold estimation in neonates. Braz J Otorhinolaryngol 2019; 85(2): 206–12.

20. Portonova G, Martynova O, Ivanitsky G. Age differences of event-related potentials in the perception of successive and spatial components of auditory information. Hum Physiol 2014; 40(1): 20–8.

21. Pena M, Werker J, Dehaene-Lambertz G. Earlier Speech Exposure Does Not Accelerate Speech Acquisition. J Neurosci 2012; 32(33): 11159–63.

22. Cone B, Whitaker R. Dynamics of infant cortical auditory evoked potentials (CAEPs) for tone and speech tokens. Int J Pediatr Otorhinolaryngol 2013; 77(7): 1162–73.

Fig. 1 Anatomical structure of the auditory pathway can be divided into a peripheral part, including the cochlea as a sensory organ, and a central part that conducts electrical potentials through the brain stem and midbrain to the primary cortical region, where it is subsequently evaluated and processed (scheme adopted and freely modified according to Graven S, Browne J. Auditory Development in the Fetus and Infant. NbInfant Nurs Rev 2008; 8(4): 187–93).

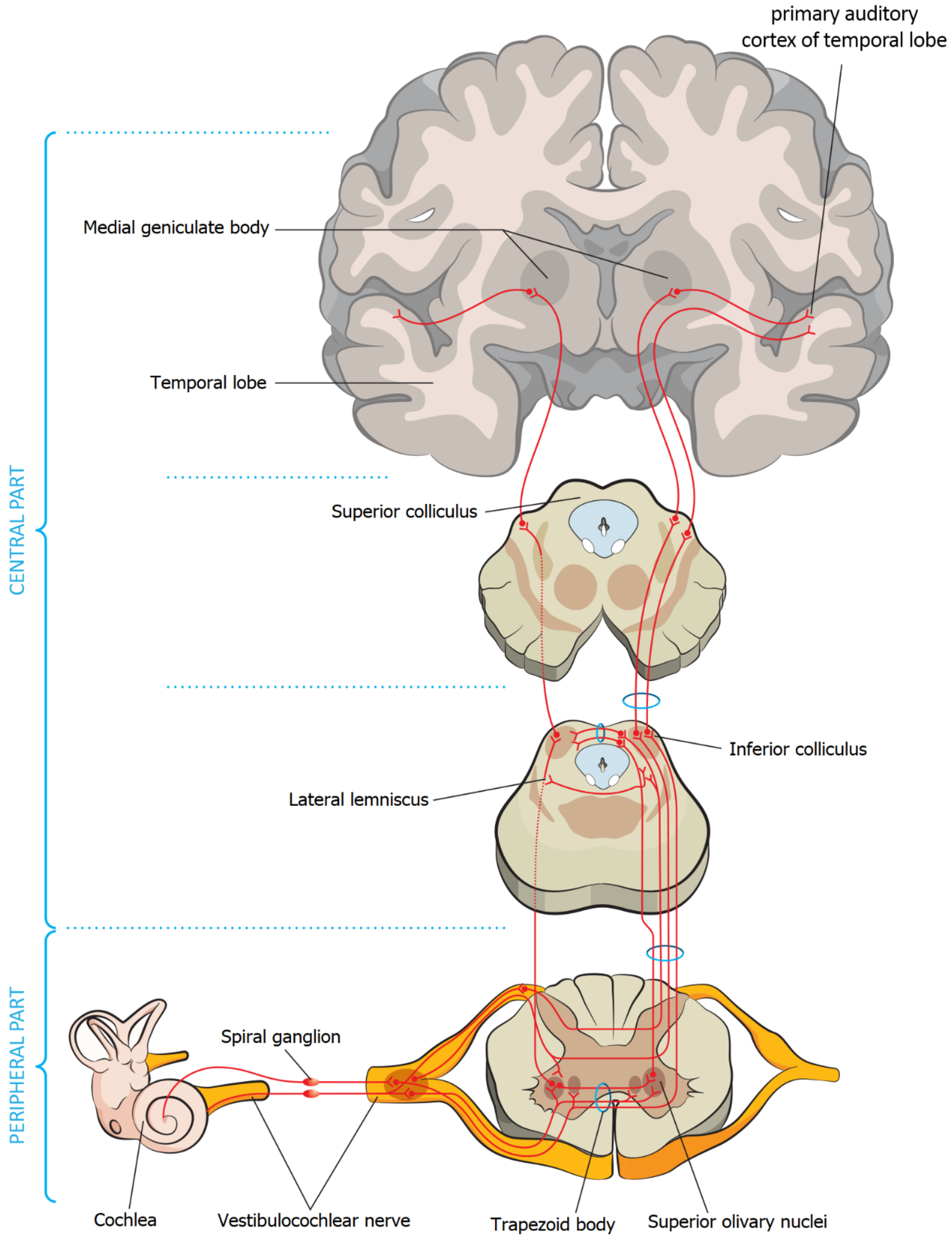


Fig. 2A–D Diagram of the gill arches and their development (marked with Roman numerals I-IV, color distribution respects the origin of tissues from individual arches also in the following figures B–D). Figures A and B also show the origin of cranial nerves important for innervation in the facial region (labeled N.V-N.X). The gill arches I and II give rise to the transmission system of the middle ear, the peripheral part of the auditory pathway. Gill arch I also develops into the tensor tympani muscle, which participates in the transmission of sound by changing the drum voltage (scheme adopted and freely modified according to Carlson, M. Nervous system. In: Human embryology and developmental biology, 5th ed. Philadelphia, PA: Elsevier/Saunders 2014: 216–45).

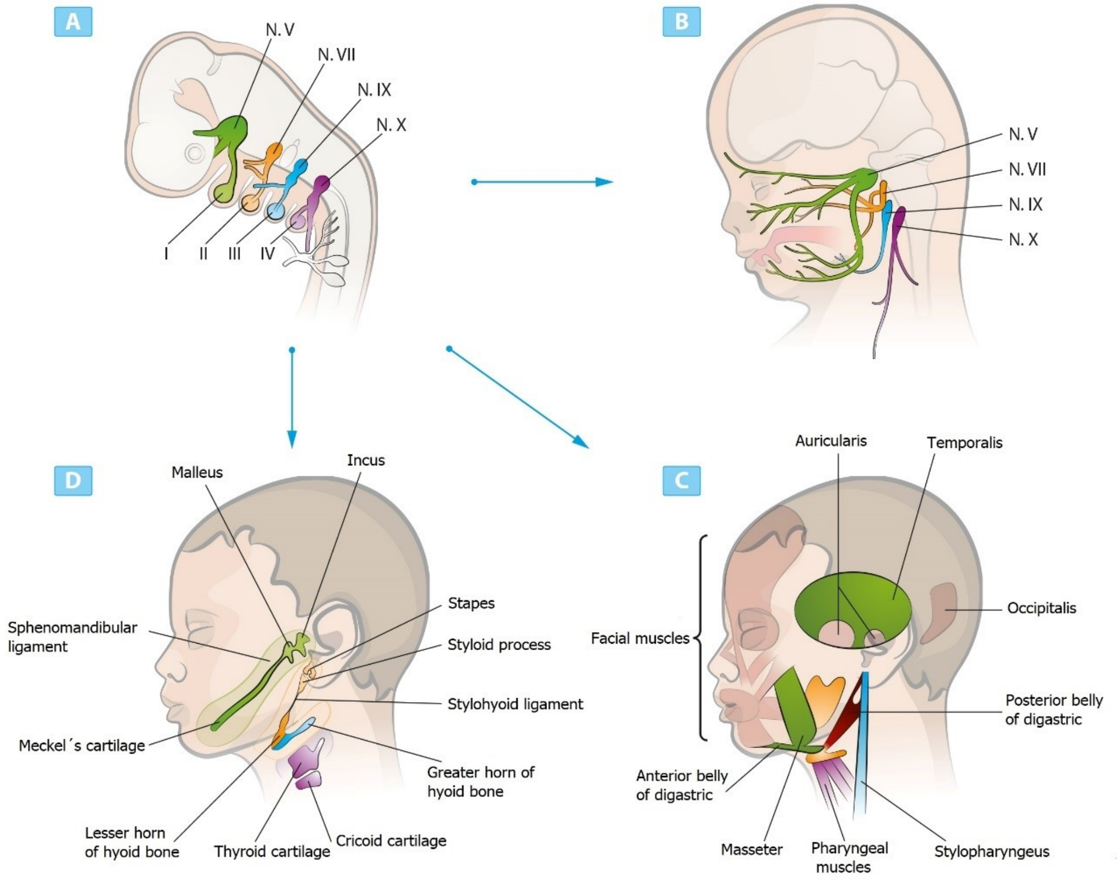
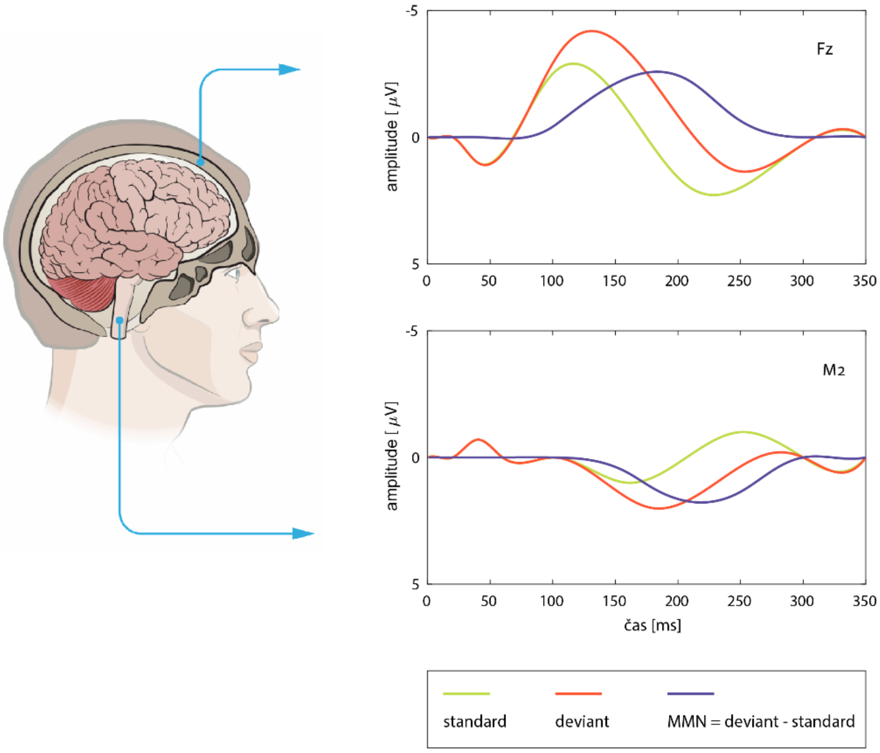


Fig. 3 Schematic representation of cortical auditory evoked potentials (CAEP) sensed by an electrode placed above the frontal area (Fz) and the processus mastoideus (M2). The frequent, standard stimulus is represented by a green curve, the rare, deviant stimulus by an orange curve. The subsequent amplitude difference of both stimuli is highlighted by a blue curve as the so-called difference wave, which peaks as mismatch negativity (MMN) at latency of about 200 ms. The amplitude of the MMN tends to be positive when measured with an electrode above the mastoid processus, in other locations it typically, in adults, has negative values (scheme adopted and freely adjusted according to Fellman V, Huotilainen M. Cortical auditory event-related potentials in newborn infants. Semin Fetal Neonatal Med 2006; 11(6): 452–8).



3. CÍLE DISERTAČNÍ PRÁCE

Základní tezí předkládané disertační práce byl předpoklad, že novorozenec již krátce po porodu dokáže adekvátně zpracovat přicházející zvukový podnět. Tato hypotéza vycházela z publikovaných studií, jejichž souhrn je uveden ve výše uvedeném přehledovém článku (viz kapitola 2.2). Hlavním cílem této práce bylo potvrzení naší hypotézy vyzrálé zvukové percepce novorozenců, tedy jejich schopnost ihned po narození rozlišit a adekvátně zpracovat zvukové kontrasty typické pro český jazyk. Vzhledem k tomu, že jsou plody před narozením stimulovány mateřským hlasem, jsme dále předpokládali, že řečové (hláskové) kontrasty vyvolají komplexnější odpověď CNS v porovnání s tónovými kontrasty. Prokázání této teze by potvrdilo, že vnímání novorozenců je pro řečové podněty již vyzrálé a do určité míry tedy podobné dospělým. Protože se sluchová dráha vyvíjí během celého těhotenství, bylo naším dalším cílem stanovení přibližné gestační hranice, od které bychom mohli tuto komplexní sluchovou percepci prokázat.

4. METODICKÉ POSTUPY ORIGINÁLNÍCH PRACÍ

K měření evokovaných potenciálů bylo využito standardního EEG záznamu pořízeného programem TruScan se vzorkovací frekvencí 3000 Hz (pilotní projekt a 1. novorozenecká studie), respektive 1000 Hz (2. novorozenecké studie). Ke zvukové stimulaci byly použity předem připravené nahrávky. Počet nahrávek zvukové stimulace se změnil po vyhodnocení 1. souboru naměřených dětí. Pro práci číslo 1 a 2 byly zvoleny řečové i tónové stimulace. Stimulace byly připraveny vždy ve dvou provedeních se změnou spektra a délky, celkem tedy 8 nahrávek - 4 pro spektrum a 4 pro délku. Pro práci číslo 3 byl zredukován počet stimulací. Byly ponechány pouze 4 řečové stimulace. Podstatou všech stimulací bylo opakování častého (tj. standardního) podnětu náhodně střídaného vzácným (tj. deviantním) stimulem za použití roving-standard paradigmatu. Pro řečovou stimulaci byly využity pro češtinu kontrastní slabiky [fa] a [fe] lišící se spektrální kvalitou samohlásek, a [fe] a [fɛ] lišící se trváním samohlásek. Počáteční souhláska [f] sloužila k lepší výslovnosti a ukotvení řečového stimulu. Slabiky v podmínce spektrální kvality byly stejné délky 220 ms. Slabiky v délkové podmínce měly trvání 180 ms a 360 ms. Délka 180 a 360 ms zastupovaly krátkou, respektive dlouhou samohlásku. Pro výzkum zpracování tónů (neřečovou stimulaci) byly zvoleny neharmonické tónové komplexy. Tyto tóny měly akustické vlastnosti pro změnu spektra i délky shodné s řečovou stimulací. Pro naši práci měly tóny 15 frekvencí. Výchozí frekvence byla 500 Hz, každá následující frekvence byla 1.5x vyšší než předcházející. Získaný signál používal 3 formanty, identické jako obsahovaly naše samohlásky [a] a [ɛ]. Délka tónových signálů byla shodná s hláskovými podněty.

4.1 METODICKÝ POSTUP PRÁCE ČÍSLO 1 („Neural Processing of Spectral and Durational Changes in Speech and Non-speech Stimuli: An MMN Study With Czech Adults“; blíže viz kapitola 5.1)

Do studie bylo zavzato celkem 32 dospělých, českých rodilých mluvčích. Nutným kritériem pro zahrnutí do výzkumné kohorty byla negativní neurologická anamnéza. Během sledování němého filmu byli tito dobrovolníci stimulování zvukovými nahrávkami pro změnu spektra i délky hlásek nebo tónů. Počet participantů stimulovaných tóny i řečovými nahrávkami byl v obou modalitách vyvážený. K měření byla použita standardní EEG čepice s 31 elektrodami, referenční elektroda byla umístěna na nose. EEG záznam byl zaznamenáván s frekvencí 3000

Hz. Studie byla schválena Etickou komisí Lékařské fakulty v Hradci Králové. Všichni participanti podepsali informovaný souhlas.

Křivky pro standardní (odvozená z odpovědí na poslední 2 časté podněty) a deviantní podnět (první odlišný podnět po předcházejícím standardním podnětu) byly získány zprůměrováním signálu ze stovek, respektive desítek prezentovaných epoch. Jejich vzájemným rozdílem byla získaná MMN, kterou lze charakterizovat jako plochu pod křivkou (area under curve, AUC) v předem definovaném okně v délce trvání 100 ms, které začalo 150 ms po změně signálu. Vypočtená AUC byla následně analyzována lineárním smíšeným modelem (linear mixed-effects model). V tomto modelu jsme zahrnuli hlavní efekty a také dvou – a třícestné interakce pro doménu (řeč, tóny), dimenzi (délka, spektrum), směr změny (periferní, centrální), lateralitu (2 kontrasty levý vs. pravý, laterální vs. středový), anterioritu (2 kontrasty centrální vs. frontální, centrální vs. parietální). Dva hlavní efekty pro anterioritu naznačují silnější (tedy negativnější) odpověď frontálně v porovnání s centrálně snímaným signálem. Centrálně snímaná odpověď byla nicméně silnější než v parietální oblasti, což potvrzovalo naši očekávanou, frontálně lokalizovanou distribuci sluchové a lingvistické odpovědi MMN. Párové srovnání dvou deviantních typů pro každou dimenzi a doménu následně nalezlo asymetrii pro spektrální kontrasty řeči - [fa] vyvolalo silnější MMN než [fɛ]. Pro všechny ostatní podmínky se křivky MMN vyvolané dvěma deviantními typy překrývaly (tj. 95 % jednoho deviantu obsahovalo průměr druhého deviantu, tedy rozdíl nebyl signifikantní při alfa 0.05). Provedení první části připravované studie ověřilo účinnost zvolených kontrastů zvukové stimulace pro český jazyk.

4.2 METODICKÝ POSTUP PRÁCE ČÍSLO 2 („Newborns´ neural processing of native vowels reveals directional asymmetries“; blíže viz kapitola 5.2)

Tato část výzkumu navázala na předchozí studii s dospělými. Novorozenecká studie byla realizována na Gynekologicko-porodnické klinice Fakultní nemocnice Hradec Králové. Jejímu zahájení předcházelo schválení Etickou komisí Lékařské fakulty v Hradci Králové. Účast novorozence byla podmíněna udělením informovaného souhlasu zákonného zástupce. Do studie bylo zahrnuto 104 donošených novorozenců, kteří pocházeli z českého jazykového prostředí. 16 dětí (z celkového počtu 120 participujících novorozenců) bylo ze studie zpětně vyloučeno. Nejčastějším důvodem byl nadměrný neklid při pořizování EEG záznamu nebo dodatečné nesplnění výběrových kritérií. Vzhledem ke zvolené oblasti zájmu a náročnosti přípravy byl redukován počet elektrod využitý k měření EEG záznamu. Referenční elektroda byla ponechána na nose, zemnící byla umístěna na tváři dítěte. Skalpové elektrody byly po rozměření přikládány jednotlivě v celkovém počtu 6 elektrod: F3, Fz, F4, C3, Cz, C4. Měření probíhalo ve spánku novorozence bez jakékoli metody tlumení. Zvuková stimulace pro řeč i tóny byla shodná jako ve výše uvedené práci 4.1. Tato stimulace byla přehrávána dítěti z reproduktorů umístěných v ušních pohárcích těsně obepínajících boltec. Intenzita zvuku byla předem nastavena v komfortní hladině tak, aby nebyla pro dítě nepříjemná a nebudila jej ze spaní (kolem 60 dB, tedy hodnoty běžného hovoru). EEG záznam byl pořizován s frekvencí 3000 Hz.

Získaná data byla zpracována lineárním smíšeným modelem (linear mixed-effects model). Naměřené amplitudy ERP a MMR byly vyhodnoceny jako AUC s danou latencí po zahájení stimulace. Tato latence byla stanovena odlišně vždy s ohledem na typ stimulace. ERP a MMN byly zprůměrovány pro každé dítě, dimensi, elektrodovou oblast a typ stimulu. Pro ERP byla zjištěna silnější odpověď u řečové stimulaci při změně délky hlásek [ɛ/ɛː] v porovnání se změnou spektra hlásek [ɛ/a]. Při vyhodnocení stimulace tóny nebyl pozorován rozdíl mezi změnou délky a spektra tónů, navíc byla měřená amplituda ERP výrazně nižší v porovnání s řečí. MMR byla získána srovnáním ERP křivek pro standardní a deviantní stimul. Opětovně byla využita AUC s latencí 80-220 ms pro časnou a 500-700 ms pro pozdní MMN po zaznění deviantní stimulace. U řečových stimulů vyvolala změna délky [fɛː] na [fɛ] více negativní MMR než opačně. Největší amplituda byla změřena nad levou hemisférou a ve střední čáře. Pro spektrální stimulace při změně hlásky [fa] na [fɛ] byla naměřena negativnější MMR než při opačné změně. Zároveň byla amplituda MMR nad levou hemisférou trvale negativní, můžeme tak mluvit o MMN. U pozdní MMN byla pouze pro změnu délky hlásek zjištěna asymetrie nad levou hemisférou.

Překvapivým závěrem této studie bylo zjištění, že novorozenci již několik dnů po narození dokáží nejen zpracovat řečové kontrasty hlásek, ale navíc pro některé modality je MMR trvale negativní. Vyhodnocená odpověď je nejlépe vyjádřena nad levou hemisférou a ve střední čáře. Index MMR pro zpracování zvukového podnětu je tedy u termínových novorozenců poměrně vyzrálý a svou morfologií již relativně podobný dospělým jedincům.

4.3 METODICKÝ POSTUP PRÁCE ČÍSLO 3 („Neural processing of speech sounds at premature and term birth: ERPs and MMR between 32 and 42 weeks of gestation“; blíže viz kapitola 5.3)

Přechozí práce prokázala schopnost donošených novorozenců zpracovávat řečové kontrasty již krátce po porodu. Cílem této práce bylo ozřejmění osvojování sluchu v perinatálním období. Jeho podstatou bylo srovnání zvukové percepce u zralých a lehce nedonošených novorozenců. Studie byla realizována na novorozeneckém oddělení Nemocnice Havlíčkův Brod. Jejímu zahájení předcházelo schválení Etickou komisí této nemocnice, účast byla podmíněna podepsáním informovaného souhlasu zákonným zástupcem. Do studie bylo zahrnuto 99 novorozenců pocházejících z českého jazykového prostředí. 3 novorozenci (z celkového počtu 102 participantů) byli ze studie dodatečně vyloučeni pro vrozenou vývojovou vadu mozku, proběhlou hypoxii a podání léků ovlivňujících funkci CNS. Novorozenci byli narozeni mezi 32. – 41. gestačním týdnem. Za hranici zralosti byl dle pediatrických zvyklostí určen gestační věk 37+0. Rozmístění elektrod bylo shodné s výše prezentovanou prací 4.2 – referenční elektroda na nose, zemnící na tváři dítěte a použity skalpové elektrody F3, Fz, F4, C3, Cz, C4. EEG záznam byl zaznamenáván s frekvencí 1000 Hz. Zvuková stimulace byla přehrávána z reproduktorů umístěných v ušních pohárcích. Stimulace byla prováděna pouze řečovými nahrávkami v obou kvalitách pro změnu spektra i délky, které byly shodné s prací 4.2.

Získaná data byla zpracována lineárním smíšeným modelem (linear mixed-effects model) jako výše uvedené studie. Amplitudy ERP a MMR řečové stimulace byly vyhodnoceny jako AUC s danou latencí po zahájení stimulace s ohledem na typ stimulace shodně s primární novorozeneckou studií. Vyhodnocením získaných ERP byla zjištěna schopnost novorozenců ve 36+1 gestačního věku spolehlivě rozpoznat české kontrastní hlásky. Od 36+6 gestačního stáří ERP odrážely rozdíly spektrálního zpracování hlásek. Rozdílné zpracování změny délky hlásek bylo zjištěno o týden později ve 37+6 gestačního věku. Ve 40+1 gestačního věku byla přítomna neurální diskriminace pro změnu délky hlásek vyjádřená negativní MMR. Na základě těchto poznatků lze tedy říci, že hranice zralosti (37+0 gestačního věku) koreluje se zjištěnou schopností novorozenců dostatečně rozlišit řečové kontrasty českého jazyka (obecně lze říci mateřské řeči). Zároveň se tím potvrzuje již prenatální vyzrávání sluchové dráhy a nezbytnost podnětů pro správný vývoj CNS plodu.

5. ORIGINÁLNÍ PUBLIKACE

5.1 Neural Processing of Spectral and Durational Changes in Speech and Non-speech Stimuli: An MMN Study with Czech Adults

Natalia Nudga[1*†], Josef Urbanec[2,3†], Zuzana Oceláková[1], Jan Kremláček[2,4] and

Kateřina Chládková[5,6]

[1] Faculty of Arts, Institute of Phonetics, Charles University, Prague, Czechia,

[2] Department of Pathological Physiology, Faculty

of Medicine in Hradec Králové, Charles University, Hradec Králové, Czechia,

[3] Pediatrics Department, Havlíčkův Brod Hospital, Havlíčkův Brod, Czechia,

[4] Department of Medical Biophysics, Faculty of Medicine in Hradec Králové, Charles

University, Hradec Králové, Czechia,

[5] Faculty of Arts, Institute of Czech Language and Theory of Communication, Charles
University, Prague, Czechia,

[6] Institute of Psychology, Czech Academy of Sciences, Brno, Czechia

*Correspondence: Natalia Nudga, nat.nudga@gmail.com

†These authors share first authorship

Neural discrimination of auditory contrasts is usually studied via the mismatch negativity (MMN) component of the event-related potentials (ERPs). In the processing of speech contrasts, the magnitude of MMN is determined by both the acoustic as well as the phonological distance between stimuli. Also, the MMN can be modulated by the order in which the stimuli are presented, thus indexing perceptual asymmetries in speech sound processing. Here we assessed the MMN elicited by two types of phonological contrasts, namely vowel quality and vowel length, assuming that both will elicit a comparably strong MMN as both are phonemic in the listeners' native language (Czech) and perceptually salient. Furthermore, we tested whether these phonemic contrasts are processed asymmetrically, and whether the asymmetries are acoustically or linguistically conditioned. The MMN elicited by the spectral change between /a/ and /+/ was comparable to the MMN elicited by the durational change between /+/ and /+:/, suggesting that both types of contrasts are perceptually important for Czech listeners. The spectral change in vowels yielded an asymmetrical pattern manifested by a larger MMN response to the change from /+/ to /a/ than from /a/ to /+/. The lack of such an asymmetry in the MMN to the same spectral change in comparable non-speech stimuli spoke against an acoustically-based explanation, indicating that it may instead have been the phonological properties of the vowels that triggered the asymmetry. The potential phonological origins of the asymmetry are discussed within the featurally underspecified lexicon (FUL) framework, and conclusions are drawn about the perceptual relevance of the place and height features for the Czech /+/-/a/ contrast.

Keywords: mismatch negativity, auditory processing, vowels, phonology, perceptual asymmetries

INTRODUCTION

Speech perception is a cognitive process which transforms the acoustic signal into respective neural representations in the human brain. One of the most fundamental properties of human speech perception is the ability to detect phonetic and phonological contrasts. Sensitivity to such contrasts has been examined by the means of behavioral tests (discrimination or categorization tasks) (Repp and Crowder, 1990; Polka and Bohn, 2003; Johnson, 2015) as well as via techniques that monitor brain activity, such as event-related potentials (ERPs) measured with electroencephalography (EEG; Eulitz and Lahiri, 2004; De Jonge and Boersma, 2015) or their magnetic equivalents measured with magnetoencephalography (Scharinger et al., 2016; Højlund et al., 2019). The most common ERP component used to study the brain response to an auditory contrast is the mismatch negativity (MMN). The MMN response is elicited by an irregularity, typically when a series of frequently presented stimuli, standards, is interrupted by a different infrequent stimulus, deviant. ERP studies show that the magnitude of the MMN reflects the extent of the perceived difference between the standard and deviant, whereby not only the acoustic distance but also the category membership of the stimuli modulate the strength of the response (Näätänen et al., 1997). The MMN can thus be used to estimate the linguistic importance and relevance of phonetic differences between stimuli for speech perception.

The auditory ERP component MMN and its magnetic correlate MMNm have been used to assess the neural processing of both vowels and consonants, and to study the relevance of qualitative, or less commonly, quantitative phonemic contrasts. Ylinen et al. (2005) studied the processing of consonant quality and quantity via MMN, focusing on stop consonants /p/, /p:/, /t/, and /t:/. In their experiment, the plosive [t:] served as the standard, [t] as a quantity deviant, [p:] as a quality deviant, and [p] as a double deviant (all embedded in the same [i_i] frame). The MMN elicited by the double deviant was approximately equal to the sum of the quantity- and quality-deviant MMNs and the authors concluded that consonant quality and quantity are processed independently. Their results also show that the quantitative change of the consonant elicited greater and earlier MMN response than the qualitative change. This finding of differential strength of processing of phoneme quality and quantity could be specific to plosive consonants. In vowels, for instance, a change in quality is much more salient than a change in plosive consonant place of articulation. The question thus remains how robustly quality versus quantity changes are processed in vowels.

Previous studies focusing on vowels show that changes in vowel spectral quality elicit a larger MMN in listeners for whom these changes represent a linguistic, i.e., phonemic change, than in listeners for whom these changes are not phonemic (Näätänen et al., 1997). Similarly, changes in the duration of vowels elicit a stronger MMN response in listeners in whose native language vowel length is phonemic than in listeners for whom it is not (Kirmse et al., 2008; Hisagi et al., 2010; Chládková et al., 2013). The effect of native phoneme inventory on both vowel quality and vowel length processing is indisputable, however, it has not yet been shown how the neural processing of vowel length and vowel spectral quality compare to one another.

The present study therefore aims to investigate and compare the neural processing of vowel duration and vowel quality of adult speakers of a language in which both vowel quality and vowel length have a contrastive role. Obtained results will also show if MMNs evoked by changes in vowel quality and quantity match with the pattern obtained by Ylinen et al. (2005) for plosive consonants, in which greater average MMN was observed in case of a quantity change.

A number of studies exploring the sensitivity to phonemic contrasts have encountered a phenomenon called perceptual asymmetry. Perceptual asymmetries can be observed when participants more readily process or respond to a change when category A is presented before category B than vice versa. Such findings imply that the perceptual space differs from the physical space and that due to its asymmetric nature its properties cannot be captured by Euclidean geometry (e.g., distances in the vowel formant space). Asymmetry in perception has been investigated for various types of stimuli including color, line orientation, numbers (Rosch, 1975), geometric figures (Tversky and Gati, 1978) as well as vowels (Polka and Bohn, 2003, 2011; Eulitz and Lahiri, 2004; De Jonge and Boersma, 2015), and consonants (Schluter et al., 2016; Cummings et al., 2017; Højlund et al., 2019). Vowel perception asymmetry has been studied by means of reaction time or accuracy in discrimination tasks, where a reversed order of stimuli led to the significant difference in the measured parameters. Asymmetrical perception of vowels has also been attested in neurolinguistic MMN studies, when the roles of standard and deviant stimuli were switched (Eulitz and Lahiri, 2004; De Jonge and Boersma, 2015). For instance, De Jonge and Boersma (2015) found asymmetrical patterns in vowel perception when comparing MMN responses of French listeners to contrasts among four French vowels [y, u, ø, o]. Their results showed that the MMN evoked by a change from a high vowel such as [u] toward a high-mid vowel such as [o], and by a change from a back vowel such as [u] to a front vowel such as [y] was significantly larger (i.e., more negative) than vice versa. In addition to the asymmetry, they found that the average MMN resulting from a change in vowel place (backness or frontness) was significantly larger compared to the MMN resulting from a change in vowel height.

There are several hypotheses and theories that offer explanation to the perceptual asymmetry phenomena. According to Repp and Crowder (1990), perceptual asymmetries are caused by different rates of memory decay, which, as the authors argued, is slower for more prototypical (or less ambiguous) vowels. They concluded that at either point of a vowel continuum the difference between stimuli is more detectable when the more salient vowel comes second in a pair, and thus serves as the subject of comparison.

Polka and Bohn (2003, 2011) proposed the natural referent vowel (NRV) framework which operates with the concept of peripheral vowels and aims to explain language-general, i.e., auditorily-based, patterns in infant speech perception. Peripherality acoustically coincides with formant focalization, that is the convergence of two formant frequencies in a vowel (Schwartz et al., 2005). In a focal vowel, the proximity of two formants strengthens their respective amplitudes and results in a perceptually prominent frequency band. According to the NRV framework, a difference is more detectable for a change from a less peripheral, or non-focal, to a more peripheral, or focal, vowel than vice versa. Along those lines, the difference between two vowels such as [u] and [y] should be more readily detectable, i.e., perceived as greater, when [y] is presented before [u] than vice versa. Note that such NRV-based asymmetry is opposite to the asymmetries obtained by De Jonge and Boersma (2015) who tested adults (and it is opposite also to the asymmetries obtained by Wanrooij et al., 2014 for infants). Although not originally proposed as an explanation for asymmetries in the neural processing of vowels, it seems viable that a more detectable difference between stimuli leads to a stronger MMN response (as shown by e.g., Näätänen et al., 1997). Therefore, the NRV can be used to formulate acoustically-based predictions for MMN such that a focal (i.e., perceptually more salient) deviant should elicit a stronger MMN than a non-focal deviant.

Repp and Crowder as well as Polka and Bohn have based their theories of vowel perception asymmetry on the acoustic properties of vowels, while other authors, namely, Lahiri and Reetz (2002) have approached this phenomenon from the phonological point of view and formulated the featurally underspecified lexicon (FUL) theory. Their theory explains the perceptual asymmetries through reference to phonological representations, postulating that a change from a stimulus specified for a particular phonological feature to a stimulus underspecified for that feature is processed more strongly than a change in the reversed order. The predictions of the FUL theory have been borne out by a number of studies (Eulitz and Lahiri, 2004; Lipski et al., 2007; Scharinger et al., 2012, 2016; De Jonge and Boersma, 2015; Schluter et al., 2016).

Considering a vowel contrast such as one between a focal and phonologically specified /a/ and a non-focal and underspecified /ɛ/, one can see that an NRV-like asymmetry predicted by acoustics (i.e., a stronger response to a change from /ɛ/ to /a/) does not necessarily coincide with an asymmetry predicted by the phonological FUL framework (i.e., a stronger response to a change from /a/ to /ɛ/). Crucially, predictions based on phonological representations can also differ depending on the adopted phonological theory. If we again consider the vowels /a/ and /ɛ/, then according to the FUL theory, /ɛ/ is underspecified for feature [LOW]. However, in Element theory (Harris and Lindsey, 1995) which describes vowels in terms of elements |A|, | I|, and | U|, it is /a/ that contains 1 element and is thus underspecified in comparison to /ɛ/ which contains 2 elements. Consequently, one could hypothesize that it is /a/ and not /+/ that should evoke greater MMN response when presented as a deviant. Although the predicted perceptual (MMN) asymmetries differ across phonological frameworks, they have been mainly tested within the FUL framework. An exception is De Jonge and Boersma (2015) who contrasted FUL and Element theory and whose MMN data from French adults supported FUL. Because it is the most widely researched phonological framework in the MMN literature, the present study adopts FUL as the basis for phonological predictions and contrasts it with NRV-like acoustic predictions.

As introduced above, the present experiment focuses on the MMN to vowel quality and vowel length contrasts which are both phonemic in the listeners' native language, Czech. The specific contrasts are /ɛ/-/a/ and /ɛ/-/ɛː/, for vowel quality and vowel length, respectively. Since spectrum can be a secondary perceptual cue to vowel length, we have selected the /ɛ/-/ɛː/ pair out of the five short-long pairs in Czech because it entails the smallest spectral difference, both in perception (Podlipský et al., 2019) and production (Paillereau and Chládková, 2019). Besides comparing the strength of the MMN elicited by the two distinct types of phonemic changes, the present experiment tests whether any MMN asymmetries exist for those vowel contrasts and if yes, whether they are phonologically or acoustically motivated.

In order to provide a further test of whether any potential asymmetries are more likely attributable to the phonology or to the acoustics, we compare Czech listeners' processing of the two vowel contrasts /ɛ/-/a/ and /ɛ/-/ɛː/ to their processing of identical acoustic differences in non-speech stimuli. The nonspeech stimuli are inharmonic tone complexes with the first three formant frequencies and duration identical to those of the vowels /a/, /ɛ/, and /ɛː/; they are thus comparably complex as the vowels but not confusable with speech. If the potential asymmetries are acoustically conditioned, they should be found in both the non-speech and the speech conditions in the present study. If, on the contrary, the asymmetries are (at least to some extent) phonologically based the pattern of results should differ across speech and non-speech.

According to Polka and Bohn (2003, 2011), the acoustic properties of our stimuli predict a greater MMN when a focal vowel (or tone complex) is the deviant and a non-focal vowel (or tone complex) is the standard. In that respect, the vowel /a/ and the /a/-like tone are focal because their first and second formants are close to one another, concentrating (focalized) energy in the F1–F2 frequency band. In contrast, the first and second formants of the vowel /ɛ/ and the /ɛ/-like tone are relatively far apart and thus contain non-focalized energy. Acoustically, the change from the non-focal /ɛ/ (-like tone) to the focal /a/ (-like tone) should elicit a stronger MMN response than a reverse change. As for the durational dimension, for which focalization has not been formally defined, intuitively a longer stimulus is more prominent than a shorter stimulus. The acoustically-motivated prediction then is that a change from the short /ɛ/ (-like tone) to the long /ɛ/ (-like tone) will elicit a greater MMN than vice versa. This direction of predicted asymmetry is further in line with previous findings that the addition of information is more detectable than its deletion (Timm et al., 2011).

The alternative, phonologically-based predictions for asymmetries are made in accordance with the featural (under)specification framework (Lahiri and Reetz, 2010), which states that the magnitude of the MMN will be greater in case of change from a fully specified vowel to an underspecified vowel than vice versa. Czech central low vowel /a/ and front mid vowel /ɛ/ differ both in the horizontal plane and in height, nevertheless from the phonological point of view there are distinguished only by means of the feature [LOW] (which is specified for /a/ but not for /ɛ/) as they are both underspecified with respect to the feature [BACK]. Therefore, in conformity with the FUL theory, we expect a greater MMN response when underspecified /ɛ/ is a deviant. Regarding the quantity contrast, according to some authors the difference between Czech /ɛ/ and /ɛ:/ lies in the feature [LONG], which is specified for /ɛ:/ (Palková, 1994, p. 206, Skarnitzl et al., 2016, p. 101). This means that in the vowel quantity condition, /ɛ/ is again underspecified, and the MMN should be larger when /ɛ/ is a deviant and /ɛ:/ is a standard.

Predictions of the vowel perception asymmetry in terms of relative magnitude of the MMN response are summarized in **Table 1**. For the complex tone stimuli, the asymmetrical behavior is expected based solely on the acoustical approach, and thus coincides with the first row of **Table 1**. To sum up, the present study has two goals. Firstly, it compares the neural processing of vowel length and vowel quality in a language that uses both types of contrasts phonemically [similarly to the comparison of consonantal quality and consonantal length reported by Ylinen et al. (2005)]. Secondly, it tests whether there are any directional asymmetries in the perception of vowel length and/or vowel quality and whether they can be explained by the vowels' acoustic properties or phonological specification.

MATERIALS AND METHODS

Stimuli

We created two sets of stimuli, one set for the speech condition and one set for the non-speech condition. The speech stimuli were naturally produced, edited consonant-vowel (CV) syllables [fɛ] and [fa]. The formants were stable throughout the vowels and corresponded to the Czech low-mid front /ɛ/ and low /a/, respectively. The first three formants of [ɛ] in [fɛ] were 755 Hz, 1646 Hz, and 2710 Hz, and the first three formants of [a] in [fa] were 864, 1287, and 2831 Hz; these values are in line with the formants of Czech vowels produced by women reported by Skarnitzl and Volín (2012). The duration of the vowels [ɛ] and [a] (extracted from the CV frames) was modified using PSOLA in Praat (Boersma and Weenink, 1992–2020). The vowel [a] had a duration of 220 ms, and [ɛ] was resynthesized with three durations, namely, 220, 180, and 360 ms, which met the following conditions: 220 ms was judged (by three expert phoneticians) as a typical duration of the mid and low short vowels in an isolated CV syllable, 360 ms represented a long vowel in a CV syllable that was not perceived as unnaturally exaggerated, and short /ɛ/ with the duration of 180 ms was considered to be sufficiently distinct from the long /ɛ:/.[1] In order to create the stimuli, we cut out the initial fricative consonant [f] from one recorded syllable and combined it with the target [a] and [ɛ] vowels, such that the fricative [f] was identical across all four speech stimuli and had a duration of 150 ms. None of the created [f] CV syllables carries lexical or morphological content in Czech. The speech stimuli had been used in a behavioral study on vowel perception with Czech-exposed infants (Paillereau et al., 2021), and recently, along with the non-speech stimuli described below, in an ERP study with Czech newborns (Chládková et al., under review).

To test the discrimination of a spectral contrast, the non-focal [fɛ] and the focal [fa] lasting for 220 ms each were used. The vowel [a] is considered focal because the distance between its first and second formant is da = 2.07 Bark, while the vowel [ɛ] in [fɛ] is non-focal because its first two formants are spread apart by d" = 4.08 Bark. The difference between [a] and [ɛ] thus lies in their perceptual prominence, where [a] is the more prominent one. The discrimination of a durational contrast was tested by the short 180-ms [fɛ] and long 360-ms [fɛ]. Similarly as for the spectral dimension, the short and the long vowel differ in their perceptual prominence, where the short one contains energy over a shorter time interval (i.e., less energy in total) as can thus be seen as perceptually less prominent stimulus than a long vowel represented by energy in a longer time interval. The intensity of the stimuli was scaled by peak to be matched across all the 4 different syllables.

The non-speech stimuli were inharmonic tone complexes with spectral and durational properties mimicking those of the vowels described above. Inharmonic tone complexes are comparably complex as vowels in that their source signal contains a series of fundamental frequency harmonics and is filtered with vocal-tract like formants. At the same time, the inharmonic tone complexes are not confusable with vowels because their source signal frequencies are spaced inharmonically (Goudbeek et al., 2009; Scharinger et al., 2014). The tone complexes in the present experiment had 15 inharmonically spaced frequency components, the first one at 500 Hz and every following being 1.15 times higher. The inharmonic source signal was filtered with three formants, namely, for the focal spectral condition with the formants of [a], for the non-focal spectral condition and the short and long durational condition with the formants of [ɛ]. Durations of the non-speech stimuli were identical to the durations of the vowels from the speech condition. The amplitude was ramped linearly over 5 ms at stimulus onset and offset. Sound intensity was scaled to be identical across all the four stimuli. As in the speech condition, the [a]-like focal tone (prominent) and the [ɛ]- like non-focal 220-ms tone (non-prominent) were used to test discrimination of spectral differences, and the 180-ms [ɛ]-

---

[1] We did not adopt the 220-ms stimulus as a short counterpart of the 360-ms /ɛ/ because the resulting long/short ratio 1.6 is more typical of the high front Czech vowel pair while for mid-low vowels the ratio is closer to 2 (Paillereau and Chládková, 2019).

like tone (non-prominent) and the 360-ms [ɛ]- like tone (prominent) were used to test discrimination of duration differences.

Presentation Paradigm

The stimuli, i.e., the individual syllables or the individual tone complexes, were presented in a roving-standard paradigm (Haenschel et al., 2005; Garrido et al., 2008; Cooper et al., 2013). Four presentation blocks were created, one for each domain (speech and non-speech) and dimension (spectrum and duration) combination. For speech spectrum, the paradigm started with 8 tokens of [fɛ] and continued with 100 trains of [fɛ] and [fa] each, alternating in series of 4–8 identical stimuli. The count of 4–8 was pseudorandom, fulfilling the condition that each count eventually occurred 20 times. The number of presented tokens was 608 for [fɛ], and 600 for [fa]; summing up to a total of 1208 stimuli in each block. Stimulus onset asynchrony was 1.09 s. Total presentation time per block was 22 min. The blocks for speech duration were created in an identical way, alternating series of short [fɛ]s and the long [fɛː]s. Analogous presentations were made for non-speech spectrum and non-speech duration. Each participant was tested with either the two speech blocks, or the two non-speech blocks. Stimulus domain thus varied between participants and dimension within participants, with the order of durational and spectral presentation counterbalanced.

Participants and Procedure

A total of 32 adult volunteers participated in the experiment. They were monolingually-raised native speakers of Czech, ages 18–28 years (mean age 24 years, 19 women, 13 men). They did not have any history of neurological or hearing disorders and reported to be right-handed.

Participants were tested in a quiet room at the Faculty of Medicine in Hradec Králové. Prior to the experiment, they filled in a demographic background questionnaire and signed an informed consent form. Half of the participants was randomly assigned to the speech condition and the other half to the nonspeech condition. Within each condition, a participant received two blocks, one presenting changes in stimulus duration and the other with changes in stimulus spectral quality; the order of the blocks was counterbalanced across participants. Between the two blocks, there was a 5-min break. During auditory stimulation, participants watched a muted movie with Czech subtitles. Participants were instructed to focus on the movie and ignore the sounds. The experiment followed the standards for research with humans and was approved by the ethics committee of the Faculty of Medicine in Hradec Králové.

Electroencephalography and ERP Processing

The EEG was recorded from thirty one Ag/AgCl electrodes Fp1, Fp2, F7, F3, Fz, F4, F8, CP4, C3, Cz, C4, TP8, FT7, P3, Pz, P4, FC3, FC4, FT8, M1, M2, OPz, AFz, P7, P8, T7, T8, CPz, FCz, TP7, CP3 referenced to an electrode placed on the nose. The EEG was recorded at a 3000-Hz sampling rate with a bandwidth of 0.3–100Hz (DEYMED Diagnostic s.r.o., Czechia). After bandpass filtering 0.2–40 Hz using EEGLab (Delorme and Makeig, 2004), the data were down-sampled to 300 Hz and epoched with MATLAB release 2020a (MathWorks, United States). The epoch started 100 ms before and ended 800 ms after the onset of the vowel or the onset of the complex tone; mean voltage of the prestimulus part (from -100 to 0 ms) was subtracted from every epoch.

Deviant waveforms were derived from every first stimulus in the row of 4–8 repeated tokens, standard waveforms were derived from the last two stimuli in the row of 4–8 repeated tokens. Standard and deviant grand-average waveforms at central channels and the MMN topographies are shown in **Figure 1**. The individual ERPs were calculated as an average of epochs with absolute amplitude under 50 µV. The ERPs were additionally digitally filtered off-line by a smoothing Savitzky-Golay filter (first polynomial order, window of 21 samples).

Difference waves were computed by subtracting the averaged standard ERP from the averaged deviant ERP elicited by physically identical stimuli, e.g., the difference waveform for the [a]-deviant was computed by subtracting the [a]-deviant ERP from the [a]-standard ERP. From the difference waves, the MMN was quantified as area under curve in a pre-defined 100-ms window that started 150 ms after change onset. The window of analysis was determined based on previously published results (Näätänen et al., 1997, 2004; Eulitz and Lahiri, 2004; De Jonge and Boersma, 2015) and visual inspection of the curves, and thus has been set 150–250 ms after vowel or tone onset for the spectral condition and 330–430 ms after vowel or tone onset for the durational condition (where the onset of change was determined as the duration of the short vowel/tone, i.e., 180 ms).

Statistical Analyses

The calculated AUC were analyzed with a linear mixed effects model (packages lme4, lmerTest in R, Bates et al., 2015; R Core Team, 2016; Kuznetsova et al., 2017). We modeled the main effects and all two- and three-way interactions of Domain (-speech, +non-speech), Dimension (-duration, +spectrum), and Deviant (-prominent, +nonprominent), as well as the main effects of Laterality (2 contrasts: -left +right, -lateral +midline) and Anteriority (2 contrasts: -central +frontal, -central +parietal). The random effects structure modeled a per-participant intercept and slopes for Dimension and Deviant.

RESULTS

The summary of the modeled fixed effects is presented in **Table 2**. As indicated by the significant intercept, overall there was a reliable MMN, estimated as -48 ±15 µV _ ms (p = 0.003). The two main effects for Anteriority suggest that the MMN was stronger (more negative) at frontal than at central sites, where it in turn was stronger than at parietal sites, thus following the expected frontally-localized distribution of the auditory and linguistic MMN response.

Regarding the predictors relevant for our research questions, there was a three-way interaction of Deviant, Dimension, and Domain[2]. To unpack the triple interaction, **Figure 2**

---

[2] A reviewer expressed concerns about a potentially low power of our experiment. We therefore simulated the power curves associated with an effect equal to the one we obtained, as well as a smaller effect, using the simr package in R (Green and MacLeod, 2016). For the simulations, we created a new model using the parameters of the initial model and calculated its power for various number of respondents for the effect of three-way interaction of Deviant, Dimension, and Domain. The obtained power curve indicated that to reach power of 80%, even with the smaller effect size (i.e., the lower bound of 95% CI of the mean estimated effect in our study) for the critical

visualizes the estimated means and confidence intervals [modeled using the R package ggeffects, Lüdecke (2018)]. Pairwise comparisons of the two deviant types on each dimension and in each domain reveal that an asymmetry between the two deviants was found in speech for the spectral contrast: [fa] elicited a stronger MMN than [f+] {[fa] mean = -95 µV × ms, CI = (-164; -27), [f+] mean = -17 µV × ms, CI = (-84; 49)}; in all other conditions the MMNs elicited by the two deviant types overlapped (i.e., the 95% CI's of one deviant contained the mean of the other deviant, which implies that the difference is not significant at alpha 0.05).

DISCUSSION

The first question addressed by this experiment was whether the neural processing of phonemic vowel quality differs from the neural processing of phonemic vowel length. To that end, we assessed the neural mismatch response (MMN) in adult speakers of Czech listening to changes between [fɛ] and [fa] and to changes between [fɛ] and [fɛ:] syllables, where both types of change represent a phonological vowel contrast. Our statistical analysis failed to detect a main effect of Dimension (or a two-way interaction of Dimension and Domain). A planned comparison of the MMN elicited by vowel quality (mean = -56 µV × ms, CI = [-111, -2]) and the MMN elicited by vowel length (mean = -44 µV × ms, CI = [-99, 11]) suggests a large overlap across the two types of vowel change, lending support to the conclusion that vowel length and vowel quality changes evoke comparable neural response in Czech adult listeners. Our results for vowels are thus different than the MMN patterns observed by Ylinen et al. (2005) for length and quality changes in plosive consonants.

If we consider the spectral and durational difference between the stimuli in just-noticeable difference units (JND), the Euclidean distance between the first three formants of the [a] and [ɛ] stimuli is equal to 5.1 JND, whereas the durational difference between the [ɛ] and [ɛ:] stimuli equals 12.8 JND [JNDs computed assuming the discrimination threshold of 0.3 bark for vowel formants, Kewley-Port (2001) and a 5 ms discrimination threshold to the reference value of 90 ms for vowel duration, Nooteboom and Doodeman (1980)]. Even though the JND in duration is more than 2 times greater than the JND in spectrum, the average MMNs elicited by each of the changes were not found to differ. Speculatively, this could be taken as an indication that the contrasts have been processed based on their phonological difference rather than the acoustic distance.

The second aim of the experiment was to test whether the vowel contrasts are processed asymmetrically, and if yes, whether the asymmetries are attributable to the acoustic or the phonological properties of the vowels. To that end, we compared the MMN elicited by changes in vowels to the MMN elicited by identical changes in non-speech stimuli. Regarding the spectral contrast, an acoustically-based approach formulated under the NRV framework (Polka and Bohn, 2003, 2011) predicted a larger MMN in case of vowel change from /ɛ/ to /a/ than vice versa. When comparing vowels /ɛ/ and /a/, the latter one is auditorily focal, or perceptually more salient, since its first and second formants are close to each other such that they merge into one prominent frequency band. In contrast, the first two formants of /ɛ/ are farther apart, resulting in vowel /ɛ/ assigned to the nonfocal, perceptually less prominent, element of the comparison. Thus, under the acoustically-based approach, we expected a larger MMN when /a/

---

three-way interaction, a total of 20 participants (i.e., 10 per group) would be sufficient. From this we conclude that our experiment with 32 participants, i.e., 16 per group, is not underpowered.

was the deviant, and smaller MMN is expected when /ɛ/ was the deviant in the present experiment. Concerning the durational difference in vowels, a long vowel, here /ɛ:/, contains acoustic energy over a longer time interval, and is thus inherently more auditorily prominent than a short vowel of the same quality, here /ɛ/. Therefore, for the change between /ɛ/ and /ɛ:/, the acoustically-based approach predicted greater MMN when the long /ɛ:/ was the deviant than when the short /ɛ/ was the deviant. Crucially, if perceptual asymmetries in vowels were acoustically conditioned, the same asymmetries were expected to be observed in the non-speech condition, which compared MMN to the changes between /ɛ/-like and /a/-like complex tones, as well as between /ɛ/-like and /ɛ:/-like complex tones. Alternatively, if any detected asymmetries did not conform to the acoustically motivated predictions, or were not detectable in the non-speech stimuli, they could be attributable to the linguistic status of the vowels. The specific phonologically-based predictions were formulated in line with the FUL (Lahiri and Reetz, 2002, 2010), and predicted an opposite direction of asymmetry due to the phonological feature specification in vowel height. Since /a/ is specified for feature [LOW] and /ɛ/ is fully underspecified, greater MMN response was expected when /ɛ/ served as deviant than vice versa. As for the durational contrast, asymmetry would be caused by feature [LONG], which is specified for /ɛ:/ but not for /ɛ/, therefore predicting greater MMN response for the short vowel /ɛ/ deviant.

The statistical model revealed a significant triple interaction of Deviant, Domain and Dimension. Pairwise comparisons of the MMN across the two directions of change (i.e., the two deviants) within each condition (i.e., for each dimension and each domain) revealed an MMN asymmetry for the spectral contrast in speech. A change from [fɛ] to [fa] elicited a stronger MMN than a change from [fa] to [fɛ] (no other asymmetries were detected). On the one hand, this result shows that a change from a non-prominent to a prominent vowel is better detectable than a reverse change, which is in line with the acoustically-motivated predictions within the NRV framework and would favor an acoustically-based explanation for the asymmetry. On the other hand, however, this asymmetry was not detected in the non-speech condition where the stimuli differed in identical acoustic parameters as did the stimuli in the speech condition. Due to its lack in the nonspeech condition, we conclude that the asymmetry that we found in the processing of the spectral vowel contrast between /a/ and /ɛ/ is specific to speech and cannot be entirely acoustically based.

Another factor suggesting that the phonologically-motivated explanation for the present MMN asymmetry is more plausible is the duration of stimulus-onset asynchrony (SOA) in our experimental paradigm. SOA was fixed at 1.09 s, which is relatively long, and therefore was more likely to tap into phonological rather than purely acoustic processing (Werker and Logan, 1985). Johnson (2015) addressed the predictions for perceptual vowel asymmetries made by the acoustic and phonological frameworks and has shown that the pattern of vowel perception asymmetry is modulated by the experimental setting. He explored perceptual asymmetries in vowels via reaction time in two discrimination tasks differing in the inter-stimulus interval (ISI), where short ISI (100 ms) implied lower-level auditory listening conditions and long ISI (700 ms) induced higher-level phonemic listening conditions. The results of Johnson's experiments indicated that the phonological underspecification model of Lahiri and Reetz (2002, FUL) accurately predicted the direction of vowel perception asymmetry in the phonemic conditions, and that in the auditory listening task this direction was reversed, and instead could be explained by the hypotheses employing acoustic characteristics of sounds. Here, we uncover an asymmetry in the processing of vowel quality but did not to detect it in a comparable non-speech condition, with a same, in Johnson's terms relatively long, ISI across the two conditions (the ISI being 730 or 910 ms depending on vowel/tone duration). It therefore appears that the asymmetry we detected for a spectral contrast in vowels is likely, at least in part, phonologically based.

However, the present asymmetry with a change from [fɛ] to [fa] eliciting a stronger MMN than vice versa, is opposite to what FUL would predict. Yet it is still possible that an underspecification account be compatible with such a finding if one considers not only the backness feature (as done in most previous MMN studies testing the FUL theory) or if one sees feature specifications as language specific. The Czech vowels /a/ and /ɛ/ do not differ only in their featural specification of height as we considered (in line with previous studies on similar vowel contrasts in other languages, e.g., /ae/ vs. /ɛ/ in Scharinger et al., 2012), but also in their featural specification of place. One could thus argue that it was the (under)specification of vowel place rather than vowel height that caused the present perceptual MMN asymmetry. The feature [FRONT] is likely specified for Czech /ɛ/ but not necessarily for Czech /a/ because in the vowel system of Czech, /a/ (along with its long counterpart) is the only low vowel does not need to be contrasted by the feature place with another low vowel quality (unlike for the mid front vowel /ɛ/ which contrasts with the mid back vowel /ɛ/). The explanation that Czech listeners responded more strongly to a mismatch in the phonological specification of vowel place than to a mismatch in the phonological specification of vowel height would also be partially in line with the results of De Jonge and Boersma (2015) who examined MMN asymmetries in French listeners. Those authors found out that the changes between French front rounded and back vowels evoked greater MMN than did the changes between high and mid-high vowels, which indicates that the horizontal difference (in place) between vowels is more salient than the vertical difference (in height).

It is possible that for the Czech /a/-/ɛ/ contrast a place mismatch is more relevant than a height mismatch, or, that both are relevant phonologically but in the case of the stimuli used here, the place mismatch overrode the height mismatch. Comparing the F1 and F2 of the vowels used in the present experiment, it can be seen that the relative distance between the first formants of [a] and [ɛ] is less (namely, 2.07 bark) than the relative distance between the second formants of [a] and [ɛ] (namely, 4.08 bark). Although phonological specification operates on discretized entities, which means that the raw acoustic distance should not matter for whether or not a phonological category contrast is perceived, MMN amplitude is modulated both by linguistic and acoustic differences between standard and deviant stimuli (e.g., Näätänen et al., 1997; Phillips et al., 2000). Therefore, the apparent prime role of underspecification of vowel place (rather than vowel height) might as well be, at least partially, driven by the fact that the change in phonological place between the /a/ and the /ɛ/ was acoustically almost twice as large as the change in phonological height (i.e., 4.08 bark versus 2.07 bark). All in all, if phonological underspecification is extended to vowel place, the present results are explainable as phonologically conditioned asymmetries.

CONCLUSION

Pre-attentive processing of changes in phonemic vowel length and vowel quality by adult Czech speakers was assessed in an ERP experiment. The neural mismatch response (MMN) elicited by a change in vowel length between /ɛ/ and /ɛː/ was comparable to the MMN elicited by a change in vowel quality between /ɛ/ and /a/, suggesting that both types of phonemic changes are equally salient to Czech speakers. For the vowel quality contrast, a perceptual asymmetry was detected where a larger MMN response was found to a change from /ɛ/ to /a/ than vice versa. No such asymmetrical pattern was observed in non-speech stimuli differing in the same acoustic parameters as the vowels, which indicated that the vowel asymmetry is more likely attributable to the vowels' linguistic status, namely phonological feature specification, than (purely) to the vowel acoustics. A stronger MMN for the vowel spectral change was elicited by a switch from /ɛ/ to /a/ than vice versa, from which we have inferred that for this Czech vowel contrast it is the feature specification for place which is primarily exploited by language users. We argued that it might have been a (language-specific) underspecification in terms of place for /a/ (rather than universal underspecification in terms of height for /ɛ/, assumed by the FUL, Lahiri and Reetz, 2002, 2010) which caused that listeners more readily detected a change from a FRONT /ɛ/ to an underspecified /a/ than vice versa.

DATA AVAILABILITY STATEMENT

The data supporting the conclusions of this article and the associated analysis scripts are available from the OSF website at https://osf.io/2849m/. The raw EEG data will be made available by the authors upon reasonable request.

ETHICS STATEMENT

The studies involving human participants were reviewed and approved by the University Hospital Hradec Králové Ethics Committee. The patients/participants provided their written informed consent to participate in this study.

AUTHOR CONTRIBUTIONS

JU, JK, and KC designed and implemented the experiment. JU and ZO performed the data collection. NN, JK, and KC processed and analyzed the data. NN wrote the manuscript with contributions and edits from KC, JK, JU, and ZO. All authors contributed to the article and approved the submitted version.

FUNDING

This project was funded by the Charles University grant Primus/17/HUM/19 and project Progres Q40/7. KC andNNwere also funded by the Czech Science Foundation grant 18-01799S.

ACKNOWLEDGMENTS

The authors thank Petr Voda for his technical support.

Conflict of Interest:
The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

REFERENCES

- Bates, D., Mächler, M., Bolker, B., and Walker, S. (2015). Fitting Linear Mixed-Effects Models Using lme4. J. Statist. Software 67, 1–48. doi: 10.18637/jss. v067.i01
- Boersma, P., and Weenink, D. (1992–2020). Praat: Doing Phonetics by Computer. Available online at: http://www.praat.org (accessed date 22 November, 2018).
- Chládková, K., Escudero, P., and Lipski, S. C. (2013). Preattentive sensitivity to vowel duration reveals native phonology and predicts learning of second-language sounds. Brain Lang. 126, 243–252. doi: 10.1016/j.bandl.2013. 05.020
- Chládková, K., Urbanec, J., Skálová, S., and Kremláček, J. Newborns' neural processing of native vowels reveals directional asymmetries. Developmental Cognitive Neuroscience 2021 Dec; 52:101023. doi: 10.1016/j.dcn.2021.101023
- Cooper, R. W., Atkinson, R. A., Clark, R. A., and Michie, P. T. (2013). Event-related potentials reveal modelling of auditory repetition in the brain. Internat. J. Psychophysiol. 88, 74–81. doi: 10.1016/j.ijpsycho.2013.02.003
- Cummings, A., Madden, J., and Hefta, K. (2017). Converging evidence for [coronal] underspecification in English-speaking adults. J. Neuroling. 44, 147–162. doi: 10.1016/j.jneuroling.2017.05.003
- De Jonge, M. J., and Boersma, P. (2015). French high-mid vowels are underspecified for height. in Proceedings of the 18th International Congress of Phonetic Sciences (Glasgow: The University of Glasgow).
- Delorme, A., and Makeig, S. (2004). EEGLAB: an open source toolbox for analysis of single-trial EEG dynamics including independent component analysis. J. Neurosci. Methods 134, 9–12. doi: 10.1016/j.jneumeth.2003. 10.009
- Eulitz, C., and Lahiri, A. (2004). Neurobiological evidence for abstract phonological representations in the mental lexicon during speech recognition. J. Cogn. Neurosci. 16, 577–583. doi: 10.1162/089892904323057308
- Garrido, M. I., Friston, K. J., Kiebel, S. J., Stephan, K. E., Baldeweg, T., and Kilner, J. M. (2008). The functional anatomy of the MMN: a DCM study of the roving paradigm. Neuroimage 42, 936–944. doi: 10.1016/j.neuroimage.2008.05.018
- Goudbeek, M., Swingley, D., and Smits, R. (2009). Supervised and Unsupervised Learning of Multidimensional Acoustic Categories. Journal of Experimental Psychology: Human Perception and Performance 35, 1913–1933. doi: 10.1037/a0015781
- Green, P., and MacLeod, C. J. (2016). SIMR: an R package for power analysis of generalized linear mixed models by simulation. Methods Ecol. Evol. 7, 493–498. doi: 10.1111/2041-210X.12504
- Haenschel, C., Vernon, D. J., Dwivedi, P., Gruzelier, J. H., and Baldeweg, T. (2005). Event-related brain potential correlates of human auditory sensory memorytrace formation. J. Neurosci. 25, 10494–10501. doi: 10.1523/jneurosci.1227-05.2005

- Harris, J., and Lindsey, G. (1995). “The elements of phonological representation,” in Frontiers of phonology: atoms, structures, derivations, eds J. Durand and F.Katamba (Harlow: Longman), 34–79.
- Hisagi, M., Shafer, V. L., Strange, W., and Sussman, E. S. (2010). Perception of a Japanese vowel length contrast by Japanese and American English listeners: behavioral and electrophysiological measures. Brain Res. 1360, 89–105. doi: 10.1016/j.brainres.2010.08.092
- Højlund, A., Line Gebauer, L.,McGregor, W. B., and Wallentin, M. (2019). Context and perceptual asymmetry effects on the mismatch negativity (MMNm) to speech sounds: an MEG study. Lang. Cogn. Neurosci. 34, 1–16. doi: 10.1080/23273798.2019.1572204
- Johnson, K. (2015). Vowel Perception Asymmetry in Auditory and Phonemic Listening. UC Berk. PhonLab Ann. Rep. 2015:11.
- Kewley-Port, D. (2001). Vowel formant discrimination II: Effects of stimulus uncertainty, consonantal context, and training. J. Acoust. Soc. Am. 110, 2141–2155. doi: 10.1121/1.1400737
- Kirmse, U., Ylinen, S., Tervaniemi, M., Vainio, M., Schröger, E., and Jacobsen, T. (2008). Modulation of the mismatch negativity (MMN) to vowel duration changes in native speakers of Finnish and German as a result of language experience. Internat. J. Psychophysiol. 67, 131–143. doi: 10.1016/j.ijpsycho.2007. 10.012
- Kuznetsova, A., Brockhoff, P. B., and Christensen, R. H. B. (2017). LmerTest Package: Tests in Linear Mixed Effects Models. J. Statist. Software 82, 1–26. doi: 10.18637/jss.v082.i13
- Lahiri, A., and Reetz, H. (2002). Underspecified recognition. Lab. Phonol. 7, 637–675. doi: 10.1515/9783110197105.2.637
- Lahiri, A., and Reetz, H. (2010). Distinctive features: Phonological underspecification in processing. J. Phonet. 38, 44–59. doi: 10.1016/j.wocn.2010.01.002
- Lipski, S. C., Lahiri, A., and Eulitz, C. (2007). Differential height specification in front vowels for German speakers and Turkish-German bilinguals: an electroencephalographic study. in Proceedings of the International Congress of Phonetic Sciences XVI, 809–812 (Saarbrücken).
- Lüdecke, D. (2018). ggeffects: Tidy Data Frames of Marginal Effects from Regression Models. J. Open Sour. Software 3:772. doi: 10.21105/joss.00772
- Näätänen, R., Lehtokoski, A., Lennes, M., Cheour, M., Huotilainen, M., Iivonen, A., et al. (1997). Language-specific phoneme representations revealed by electric and magnetic brain responses. Nature 385, 432–434. doi: 10.1038/385432a0
- Näätänen, R., Pakarinen, S., Rinne, T., and Takegata, R. (2004). The mismatch negativity (MMN): towards the optimal paradigm. Clin. Neurophysiol. 115, 140–144. doi: 10.1016/j.clinph.2003.04.001
- Nooteboom, S. G., and Doodeman, G. J. N. (1980). Production and perception of vowel length in spoken sentences. J. Acoust. Soc. Am. 67, 276–287. doi: 10.1121/1.383737
- Paillereau, N., and Chládková, K. (2019). Spectral and temporal characteristics of Czech vowels in spontaneous speech. AUC Philologica 2019:19. doi: 10.14712/24646830.2019.19
- Paillereau, N., Podlipský, V. J., Šimáčková, Š, Smolík, F., Oceláková, Z., and Chládková, K. (2021). Perceptual sensitivity to vowel quality and vowel length in the first year of life. JASA Exp. Lett. 1:025202. doi: 10.1121/10.0003369
- Palková, Z. (1994). Fonetika a fonologie češtiny. Prague: Karolinum.

- Phillips, C., Pellathy, T., Marantz, A., Yellin, E., Wexler, K., Poeppel, D., et al. (2000). Auditory cortex accesses phonological categories: an MEG mismatch study. J. Cogn. Neurosci. 12, 1038–1055. doi: 10.1162/08989290051137567
- Podlipský, V. J., Chládková, K., and Šimáˇcková, Š (2019). Spectrum as a perceptual cue to vowel length in Czech, a quantity language. J. Acoust. Soc. Am. 146, EL352–EL357.
- Polka, L., and Bohn, O. S. (2003). Asymmetries in vowel perception. Speech Comm. 41, 221–231. doi: 10.1016/S0167-6393(02)00105-X
- Polka, L., and Bohn, O. S. (2011). Natural Referent Vowel (NRV) framework: An emerging view of early phonetic development. J. Phonet. 39, 467–478. doi: 10.1016/j.wocn.2010.08.007
- R Core Team. (2016). R: a language and environment for statistical computing. R Foundation for Statistical Computing. Vienna: R Core Team.
- Repp, B. H., and Crowder, R. G. (1990). Stimulus order effects in vowel discrimination. J. Acoust. Soc. Am. 88, 2080–2090. doi: 10.1121/1.400105
- Rosch, E. (1975). Cognitive reference points. Cogn. Psychol. 7, 532–547. doi: 10.1016/0010-0285(75)90021-3
- Scharinger, M., Herrmann, B., Nierhaus, T., and Obleser, J. (2014). Simultaneous EEG-fMRI brain signatures of auditory cue utilization. Front. Neurosci. 8:137. doi: 10.3389/fnins.2014.00137
- Scharinger, M., Monahan, P. J., and Idsardi, W. J. (2012). Asymmetries in the Processing of Vowel Height. J. Speech Lang. Hear. Res. 55, 903–918. doi: 10.1044/1092-4388(2011/11-0065)
- Scharinger, M., Monahan, P. J., and Idsardi, W. J. (2016). Linguistic category structure influences early auditory processing: Converging evidence from mismatch responses and cortical oscillations. NeuroImage 128, 293–301. doi: 10.1016/j.neuroimage.2016.01.003
- Schluter, K., Politzer-Ahles, S., and Almeida, D. (2016). No place for /h/: an ERP investigation of English fricative place features. Lang. Cogn. Neurosci. 31, 728–740. doi: 10.1080/23273798.2016.1151058
- Schwartz, J.-L., Abry, C., Boë, L.-J., and Vallée, N. (2005). The dispersionfocalization theory of sound systems. J. Acoust. Soc. Am. 117:4. doi: 10.1121/1.4786487
- Skarnitzl, R., and Volín, J. (2012). Referenční hodnoty vokalických formantů pro mladé dospělé mluvčí standardní češtiny. (Reference values of vowel formants of young adult speakers of standard Czech.). Akustické listy 18, 7–11.
- Skarnitzl, R., Šturm, P., and Volín, J. (2016). Zvuková báze řečové komunikace. Prague: Karolinum.
- Timm, J., Weise, A., Grimm, S., and Schröger, E. (2011). An asymmetry in the automatic detection of the presence or absence of a frequency modulation within a tone: a mismatch negativity study. Front. Psychol. 2:189. doi: 10.3389/fpsyg.2011.00189
- Tversky, A., and Gati, I. (1978). "Studies of similarity," in Cognition and Categorization, eds E. Rosch and B. B. Lloyd (Lawrence: Erlbaum).
- Wanrooij, K., Boersma, P., and van Zuijen, T. L. (2014). Distributional Vowel Training Is Less Effective for Adults than for Infants. A Study Using the Mismatch Response. PLoS One 9:e109806. doi: 10.1371/journal.pone.0109806
- Werker, J. F., and Logan, J. S. (1985). Cross-language evidence for three factors in speech perception. Percept. Psychophy. 37, 35–44. doi: 10.3758/bf03207136

- Ylinen, S., Huotilainen, M., andNäätänen, R. (2005). Phoneme quality and quantity are processed independently in the human brain. Neuroreport 16, 1857–1860. doi: 10.1097/01.wnr.0000185959.11465.9b

Table 1. Acoustically- and phonologically-based predictions of relative magnitude of the MMN response to the experimental stimuli.

	Direction of the MMN asymmetry		
Acoustics (NRV)	[ɛ] →[a]	>	[a] →[ɛ]
Phonology (FUL)	[ɛ] →[ɛ:]	<	[ɛ:] →[ɛ]

Table 2. Fixed-effects summary of the model outcomes.

Predictor	**Estimate**	**SE**	**df**	***t***	***p***
Intercept	**-47.999**	**15.150**	**31.738**	**-3.168**	**0.003**
Deviant (-prominent +non-prominent)	12.534	27.848	31.802	0.450	0.656
Dimension (-duration +spectrum)	-31.456	26.836	31.281	-1.172	0.250
Domain (-speech +tone)	4.757	30.299	31.738	0.157	0.876
Laterality (-left +right)	8.084	10.354	1057.792	0.781	0.435
Laterality (-lateral +midline)	-19.745	11.956	1057.792	-1.652	0.099
Anteriority (-central +frontal	**-46.064**	**11.956**	**1057.792**	**-3.853**	**<0.001**
Anteriority (+central +parietal)	**30.782**	**11.956**	**1057.792**	**2.575**	**0.010**
Deviant × Dimension	17.550	17.138	1068.149	1.024	0.306
Deviant × Domain	-18.349	55.695	31.802	-0.329	0.744
Dimension × Domain	-38.804	53.672	31.281	-0.723	0.475
Deviant × Dimension × Domain	**-189.978**	**34.275**	**1068.149**	**-5.543**	**<0.001**

Rows marked in bold indicate the effects with $p < 0.05$.

Figure 1. Standard and deviant grand-average waveforms at central channels (averaged cross C3, Cz, and C4), and the MMN topographies (displaying the area under curve, AUC, measured in the shaded time windows from deviant-standard differences), per Domain, Dimension, and Deviant type (arrows mark tones/vowels onset).

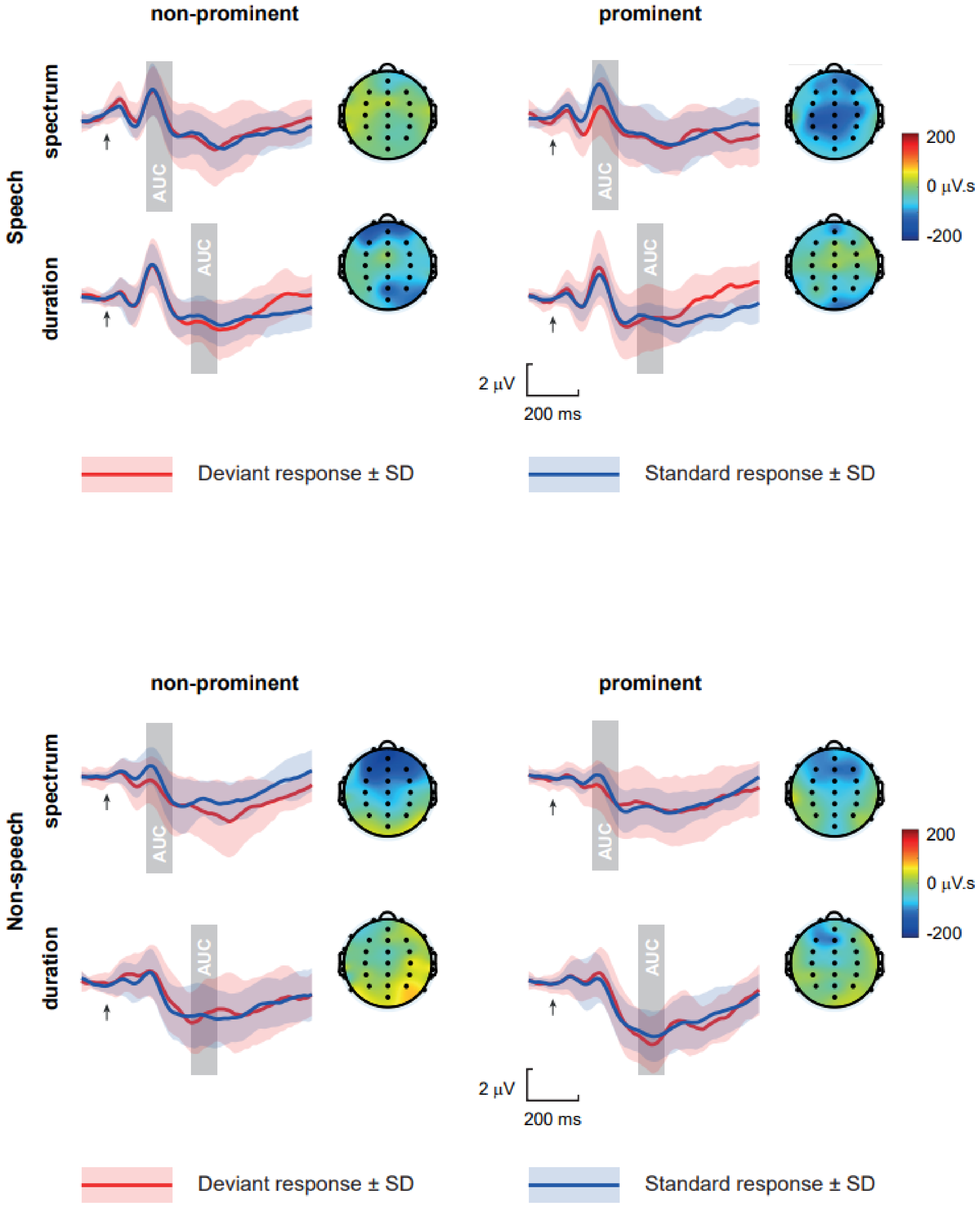
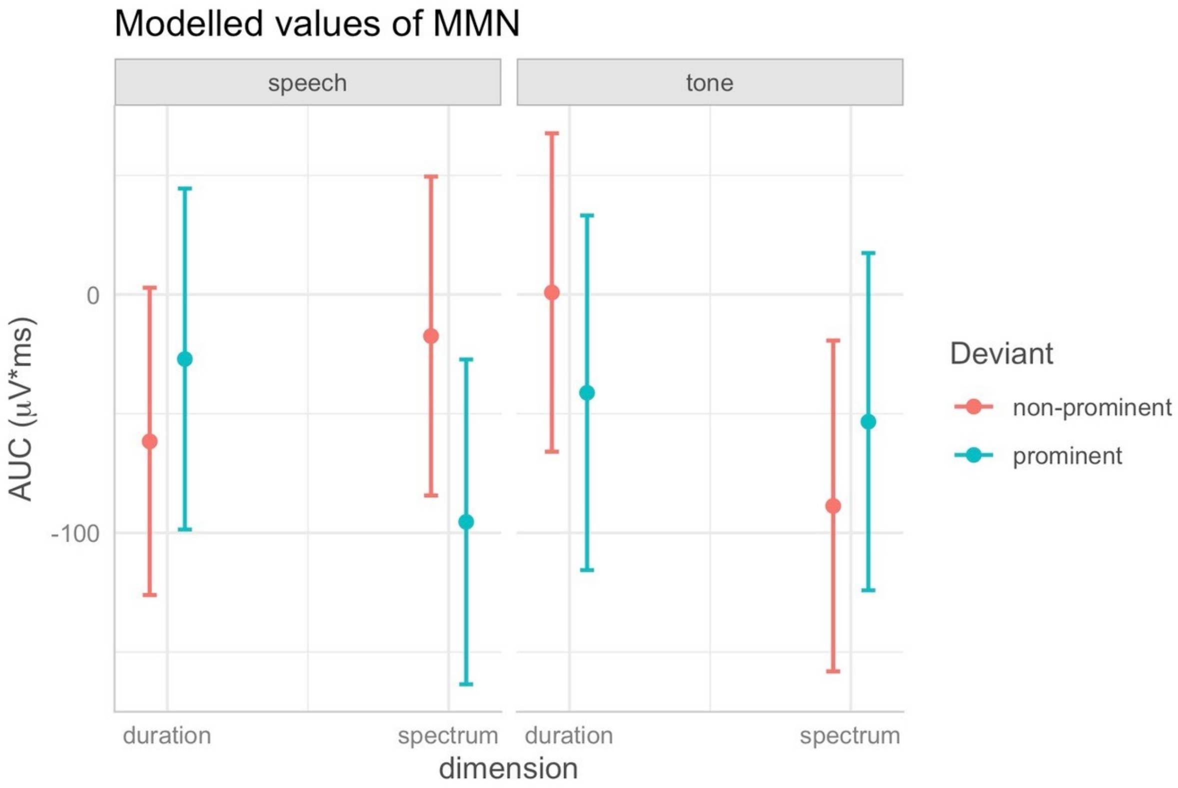


Figure 2. Unpacking the significant three-way interaction of Deviant, Dimension, and domain. The figure shows model-estimated means and 95% confidence intervals for the MMN elicited by acoustically prominent and non-prominent deviants on each dimension, separately in speech and non-speech stimuli.



5.2 Newborns´ neural processing of native vowels reveals directional asymmetries

Kateřina Chládková [a,b,*], Josef Urbanec [c,d], Sylva Skalová [e], Jan Kremláček [c,f]

[a] Institute of Czech Language and Theory of Communication, Faculty of Arts, Charles University, Nam. Jana Palacha 2, 116 38 Praha, Czechia

[b] Institute of Psychology, Czech Academy of Sciences, Hybernská 8, 110 00 Praha, Czechia

[c] Department of Pathological Physiology, Faculty of Medicine in Hradec Králové, Charles University, Šimkova 870, 500 03 Hradec Králové, Czechia

[d] Paediatrics Department, Havlíčkův Brod Hospital, Husova 2624, 580 01 Havlíčkův Brod, Czechia

[e] Paediatrics Department of University Hospital, Sokolská 581, 500 05 Hradec Králové, Czechia

[f] Department of Medical Biophysics, Medical faculty in Hradec Králové, Charles University, Šimkova 870, 500 03 Hradec Králové, Czechia

* Corresponding author at: Institute of Psychology, Czech Academy of Sciences, Hybernská 8, 110 00 Praha, Czechia. E-mail address: chladkova@praha.psu.cas.cz (K. Chládková).

Prenatal learning of speech rhythm and melody is well documented. Much less is known about the earliest acquisition of segmental speech categories. We tested whether newborn infants perceive native vowels, but not nonspeech sounds, through some existing (proto-)categories, and whether they do so more robustly for some vowels than for others. Sensory event-related potentials (ERP), and mismatch responses (MMR), were obtained from 104 neonates acquiring Czech. The ERPs elicited by vowels were larger than the ERPs to nonspeech sounds, and reflected the differences between the individual vowel categories. The MMRs to changes in vowels but not in nonspeech sounds revealed left-lateralized asymmetrical processing patterns: a change from a focal [a] to a nonfocal [ɛ], and the change from short [ɛ] to long [ɛ:] elicited more negative MMR responses than reverse changes. Contrary to predictions, we did not find evidence of a developmental advantage for vowel length contrasts (supposedly most readily available in utero) over vowel quality contrasts (supposedly less salient in utero). An explanation for these asymmetries in terms of differential degree of prior phonetic warping of speech sounds is proposed. Future studies with newborns with different language backgrounds should test whether the prenatal learning scenario proposed here is plausible.

1. Introduction

Humans learn about their mother’s voice, language, and frequently recited rhymes while still in the womb (Mehler et al., 1988; DeCasper et al., 1994; Kisilevsky et al., 2009). These early linguistic abilities have been attributed to fetal sensitivity to language prosody, that is, its rhythm and intonation (Moon et al. 1993; Granier-Deferre et al., 2011; Abboub et al. 2016). Newborn cortices indeed show specialization for listening to streams of speech over nonspeech, and process native-language speech differently from non-native speech (May et al., 2018; Sato et al., 2012).

Besides prosody, languages differ vastly in the speech segments that they employ to construct and contrast words: for instance, British English contrasts 44 segmental categories, while Central Rotokas, a language spoken in Papua New Guinea, only has 11 (Maddieson, 1986). Unlike prosody, however, whose prenatal acquisition has been studied relatively widely, the earliest linguistic development of individual speech segments is less documented. The earliest stage of segmental speech sound processing and learning is examined in the present study. We ask whether newborn infants' processing of speech sound contrasts displays any evidence of prior, i.e. prenatal, experience with those contrasts.

A review of existing literature suggests that the intrauterine linguistic development could comprise learning even of segmental properties of speech. Firstly, the speech signal in utero preserves some of the acoustic properties that cue segmental identity. Sounds' spectral properties are relatively well preserved in the range up to ~ 1000 Hz with higher frequencies being progressively attenuated by about 6 dB/octave, although these values vary across studies (see Granier-Deferre et al., 2011). The higher frequency range thus gets diminished while lower frequencies, including durational modulations, reach the fetus in a virtually unchanged form, or might even be perceptually strengthened (Richards et al., 1992). The preservation of low-frequency and durational characteristics then enables the fetus to hear and learn the language-specific intonational and rhythmic patterns (Querleu et al., 1988; Granier-Deferre et al., 2011; Abboub et al., 2016). Crucially, not only rhythm but also some segmental categories of speech are cued by frequency information below ~1000 Hz and by duration, which leads to the hypothesis that the developing human could start acquiring segmental speech categories during the prenatal period.

A normally developing fetus is able to hear and process the encountered acoustic signal. At around 20 weeks of gestation, neuronal connections in the peripheral and central auditory system begin to be formed and tonotopic organization develops in the cochlea, and from about gestational week 28 in the temporal cortex (Graven and Browne, 2008). From at least the 35th gestational week fetuses perceptually discriminate tones with frequencies 250 Hz versus 500 Hz, and vowels [i] versus [a] embedded either in a [b_] or a [b_b_] context (Shahidullah and Hepper, 1994; Lecanuet et al., 1987). However, 36-week old fetuses do not discriminate the consonantal [da]-[ta] distinction (mainly distinguished by frication above 2000 Hz) although pre-term infants born at 29–32 weeks do discriminate a (different) consonantal contrast [ba]-[ga] (Weikum et al., 2012; Mahmoudzadeh et al., 2013). These studies suggest that some speech sounds, such as consonantal contrasts cued in a high-frequency range, may not be audible and/or discriminable in utero to the same extent as some vowels or tones are.

A handful of relatively recent studies indicate that fetuses can engage in the process of speech sound learning. Partanen et al. (2013) found that infants who received prenatal training with rare pitch and vowel quality variations in a frequently exposed pseudoword [tatata] had enhanced neural processing of pitch differences at birth as compared to a group of untrained infants. Specifically, infants were more sensitive to changes in vowel fundamental frequency (averaging around 170 Hz) if they were exposed to them prenatally (Partanen et al., 2013). Besides such prenatal controlled exposure, another study suggests that newborn speech sound perception may be influenced by natural language environment. Moon et al. (2013) showed that 1- to 4-days old American English and Swedish infants differ in how they behaviourally react to American English /i/ and Swedish /y/, acoustically differentiated in the low frequency range at about 250 Hz, as well as in the higher frequency range 2–3 kHz. Infants from either group were perceptually more sensitive to variants of the non-native vowel category (in line with language-specific categorical perception), meaning that they processed native and non-native vowels differently. A reanalysis of Moon et al.'s (2013) data reported by Zhao et al. (2011) further supports the role of native language exposure during prenatal development. The native-language effect seems to have been driven by those newborns who had an older sibling (4 years or younger) – and thus likely overheard infant-directed, i.e., exaggerated and affective, speech during their prenatal development – than in infants without such a sibling. Moon et al.'s (2013) data thus indicate that the learning of native vowel categories from exposure might start already before birth.

In sum, humans can hear and are capable of learning the speech sounds of their native language before birth. Since vowels are (prenatally) the most perceptually salient sounds, they are also the focus of the present study. Languages commonly contrast anywhere between 5 and 35 vowel categories, such that within the class of vowels one will likely find various patterns and onsets of learning. In some languages (e.g. Finnish, Japanese, or Czech), duration cues not only prosody but also segmental short-long vowel contrasts. Given the veridical transmission of the durational cues to the womb, as opposed to the modulations affecting vowel spectrum, one could hypothesize that in languages with contrastive vowel length, durationally-cued vowel categories will have a developmental advantage over spectrally-cued ones. To test that hypothesis, this study focuses on two types of vowel contrasts: one durational and one spectral.

We assess the neural processing of speech sounds in one-to-three days old infants, who had been exposed to a language that systematically differentiates vowels both by duration and by spectral quality (namely, Czech). The newborns are tested on their processing of durational and spectral changes in two sets of stimuli: speech and nonspeech. Both stimulus sets contain similar acoustic patterns but in different contexts – in one context these patterns occur in vowel stimuli that specify the native-language categorical contrasts /ɛ/-/a/ and /ɛ/-/ɛː/ and in the other context they occur in complex inharmonic tones that are not interpretable as speech.

To measure whether the newborns employ categorical 'knowledge' during stimulus processing, we assess their mismatch responses (MMR). The MMR is particularly suited as an index of higher perceptual processing because it quantifies the conflict between a prediction created on the basis of one stimulus and its violation caused by another stimulus (Näätänen, 2001; Winkler & Czigler, 2012). In infants and children, the MMR has been employed to assess the formation of language-specific speech sound representations (Cheour et al., 2002; Cheour et al., 1998; Nenonen et al., 2005). Initially in development, the size of the MMR seems mainly correlated with acoustic distance between speech stimuli, but as linguistic representations come to be formed, the categorical mental encoding overrides the acoustic distance effect and becomes the primary modulator of the MMR (Cheour et al., 1998). Besides its size, the polarity of the MMR to speech has been shown to reflect the developmental stage of an individual and/or of a particular linguistic contrast, where a negative deflection of the MMR characterizes a more mature response than a positive deflection (Maurer et al., 2003; Mueller et al. 2012, Thiede et al., 2019) and/or a contrast that is easier to discriminate (Peter et al., 2016). The MMR thus seems ideal means for uncovering the extent to which newborn infants employ prior experience with speech sounds when processing different types of stimuli.

With respect to our hypothesis of developmental advantage of vowel length over vowel quality, we can formulate predictions both about the strength and the polarity of the MMR. Firstly, we expect the MMR to changes in vowel duration to be more robust, i.e. of greater amplitude than the MMR to changes in vowel spectral quality. Regarding the polarity, vowel length changes could result in a negative-going MMR while vowel quality changes in a positive-going MMR.

Studies on perceptual discrimination of vowels, with both infants or adults, often report directional asymmetries. For instance, within the /i/-/ɛ/ contrast, young ‘pre-linguistic’ infants might be more sensitive to a change from /ɛ/ to /i/ than to a change from /i/ to /ɛ/ (Polka and Bohn, 2011). Peripheral vowels like /i/ or /a/ are characterized by stable articulatory-acoustic relations, as well as by a concentration of acoustic energy in a particular frequency range (i.e. focalization), while non-peripheral vowels like /ɛ/ are not: these differential phonetic properties have been argued to cause the asymmetries in infants’ vowel perception (Polka and Bohn, 2003, 2011; Schwartz et al., 2005). Note however that not all studies with infants found such perceptual asymmetries (Wanrooij et al. 2014) and that adults may even display reverse asymmetries (Scharinger et al., 2011; Lahiri and Reetz, 2010). To account for the possibility that also newborn infants have a perceptual asymmetry, the present study employs a stimulation paradigm that allows to assess the MMR to changes in both directions within individual participants in a reasonable amount of time. No specific a priori predictions were formulated about the directional asymmetries, but they will be returned to in the Discussion.

Prior to analysing MMR, we will assess the newborns’ primary sensory responses (ERPs) to the different auditory stimuli. Physically different stimuli typically elicit different sensory ERPs, e.g. in adults the amplitude of the ERP approximately 100 ms after stimulus onset, the N1, is inversely related to vowel first formant (Scharinger et al. 2011). Since the infants tested here have normally developing hearing we predict that they will process the acoustic differences between the [ɛ] and [a]-like stimuli and between the short and long stimuli in both the speech and the nonspeech condition. Therefore, the ERPs elicited by [ɛ](-like) and [a](-like) and by short and long sounds are predicted be different[3].

To summarize, the experiment reported here investigates whether the acquisition of native vowels is underway already before birth and whether durational contrasts have an early advantage over spectral contrasts. Given the loudness and intrauterine availability of at least some vowel cues, it is likely that normally developing infants will have already started the process of category formation for the vowels of their native language. Considering the absolute veridical transmission of acoustic duration and the gradual attenuation of frequency information, we predict that durationally-cued vowel categories are at birth acquired more robustly than spectrally-cued vowel categories. Possibly, one or both types of vowel contrasts may result in asymmetric patterns in the MMR with one direction of change causing a stronger MMR response than the other direction. If the effects that we predict for vowels (the advantage of vowel length over vowel quality and/or any directional asymmetries) are due to prior exposure to the sounds they should not be observed for non-linguistic stimuli.

2. Method

2.1. Stimuli

2.1.1. Speech and non-speech segments

---

[3] 1 In fact, as the MMR is an ERP difference, reporting sensory responses should be mandatory even in adult studies because the same measured MMR may result from different conditions. For example, no response to deviant and a small response to standard will create an difference response, which might be wrongly interpreted in the sense of the prediction error even though the neural system does not respond to one of the stimuli at all (see Kremláček et al., 2016).

Speech stimuli were naturally produced, edited consonant-vowel (CV) syllables [fɛ] and [fa]. The vowel formants were stable throughout and representative of the Czech low-mid front /ɛ/ and low /a/, respectively. The first three formants (i.e. F1, F2, and F3) of [ɛ] in [fɛ] were 755 Hz, 1646 Hz, and 2710 Hz. The first three formants of [a] in [fa] were 864 Hz, 1287 Hz, and 2831 Hz. The vowels [ɛ] and [a] were extracted and their durations edited using PSOLA in Praat (Boersma and Weenink, 1992–2020). We made one [a] with a duration of 220 ms, and three [ɛ]'s, namely, 220 ms, 180 ms, and 360 ms. These durations fulfilled the following criteria: 220 ms was judged (by 3 expert phoneticians) as a typical duration of the mid and low short Czech vowels in an isolated CV syllable, 360 ms was representative of a long Czech vowel in a CV syllable that was not perceived as unnaturally exaggerated, and 180 ms was judged as sufficiently distinct from the long vowel, also based on the previously reported finding that short low and mid vowels are in Czech about half the duration of their long counterparts (Paillereau and Chladková, 2019). Note that in Czech both short and long vowels are legitimate in open syllables.

From a different recorded syllable [fɛ] we cut out the initial fricative [f], which had a duration of 150 ms, and spliced it onto the target [a] and [ɛ] vowels. The fricative [f] was thus identical across all four speech stimuli. Neither of the [f]+vowel monosyllables carries lexical or morphological content in Czech.

The four speech stimuli are visualized in Fig. 1, box I. The 220-ms [fɛ] and the 220-ms [fa] tested discrimination of a spectral contrast, which is why they are referred to as spectrally nonfocal and spectrally focal, respectively. The [a] in [fa] is focal because its first two formants (visible in the spectrograms of Fig. 1 as black horizontal bars) are close to one another (merging into a single black horizontal bar in the spectrogram); the [ɛ] in [fɛ] is termed as nonfocal, because its first and second formant are spread apart (and clearly visible as two separate horizontal bars in the spectrogram). The 180-ms [fɛ] and the 360-ms [fɛ] were used to test discrimination of a durational contrast, and are referred to as short and long, respectively. Average stimulus intensity was equated across all four syllables.

Nonspeech stimuli were inharmonic tone complexes with spectral and durational properties mimicking those of the vowels described above. Inharmonic tone complexes have a similar acoustic structure to vowels in that their source signal contains a series of exponentially spaced frequencies, and is filtered with vocal-tract like formants. At the same time, the inharmonic tone complexes are not confusable with vowels because their source-signal frequencies are spaced inharmonically (Goudbeek et al., 2009; Scharinger et al., 2014). The difference in language-likeness between the conditions was further strengthened by using CV syllables as the speech stimuli but isolated individual tone complexes as the non-speech stimuli.

The tone complexes in the present experiment had 15 inharmonically spaced frequency components, the first one at 500 Hz and every following being 1.15 times higher. The inharmonic source signal was filtered with three formants, namely, for the focal spectral condition with the formants of [a], for the nonfocal spectral condition and the short and long durational condition with the formants of [ɛ]. The tone complexes were acoustically somewhat simpler in spectral content than the vowels because they were filtered with 3 formants, while the vowels also had spectral content in higher frequencies (as can be seen in Fig. 1). Since monophthongal low vowels, such as the [ɛ] and [a] used here, are sufficiently differentiated by the first two formants (and F3 helps to normalize for talker variation, Monahan and Idsardi, 2010), the non-speech synthesis with F1, F2, and F3 was considered adequate for comparing the discrimination of vowel(-like) spectral quality across speech and non-speech. Durations of the nonspeech stimuli were identical to the durations of the vowels from the speech condition.

The amplitude was ramped linearly over 5 ms at stimulus onset and offset (in contrast to the speech stimuli, the non-speech stimuli had a more uniform amplitude envelope, as seen in Fig. 1). Average sound intensity was equated across all the four nonspeech stimuli, as well as across speech and nonspeech.

The nonspeech stimuli are plotted in Fig. 1, box II. As in the speech condition, the [a]-like focal tone and the [ɛ]-like nonfocal 220-ms tone were used to test discrimination of spectral differences, and the 180-ms [ɛ]-like tone and the 360-ms [ɛ]-like tone were used to test discrimination of duration differences. The stimuli are the same as those used in Nudga et al., 2021 who measured MMN to vowel and nonspeech contrasts with Czech adults.

2.1.2. Stimulus presentation

Stimuli were presented in a roving-standard paradigm (e.g. Haenschel et al., 2005). Four presentation blocks were created, one for each domain (speech and nonspeech) and dimension (spectrum and duration) combination. For speech spectrum, the paradigm started with 8 tokens of [fɛ] and continued with 100 trains of [fɛ] and [fa] each, alternating in series' of 4–8 identical stimuli. The count of 4–8 was pseudorandom, fulfilling the condition that each count eventually occurred 20 times. The number of presented tokens was 608 for [fɛ], and 600 for [fa]; summing up to a total of 1208 stimuli in each block. Stimulus-onset asynchrony was 1.09 s. Total presentation time per block was 22 min. The blocks for speech duration were created in an identical way, alternating series' of short [fɛ]s and the long [fɛː]s. Analogous presentations were made for nonspeech spectrum and nonspeech duration.

An individual infant was tested with either the two speech blocks, or the two nonspeech blocks. Stimulus domain thus varied between participants and dimension within participants, with the order of durational and spectral presentation counterbalanced between infants.

2.2. Participants

The participants were 104 full-term, healthy infants (16 additional infants were tested but excluded due to fussiness or noisy recording).2 Their physiological details are given in Table 1. All infants' Apgar score (vitality index) at the 10th minute after birth was 10 (highest value), and all passed the neonatal hearing test. Physiological vaginal and uncomplicated caesarean births were included. All mothers were monolingual native speakers of Czech. The infants were judged as low-risk regarding developmental language or speech-related disorders (based on absence of symptoms in parents and siblings).

2.3. Procedure

The experiment was approved by the ethics committee of the Faculty of Medicine and University Hospital in Hradec Králové, Charles University. Mothers of newly born infants who volunteered to participate did so after providing an informed consent. They received a small gift for their participation.

The experiment was administered in a quiet room at the maternity ward of the University Hospital in Hradec Králové. During the experiment, infants were asleep, lying supine in their cot (note that sleep does not seem to affect MMR in newborns, unlike in adults, Martynova et al., 2003). Auditory stimulation was through ER-3C earplugs (Etymotic research, Inc.), fitted in disposable earphones (Flexicouplers by Natus Europe, GmbH), at 67 dB SPL. If during the experiment an infant showed signs of waking up, the mother, who was present in the room throughout, was asked to calm them back to sleep. If an infant did not sleep, the experiment was terminated (this happened for 3 infants).

2.4. EEG recording and ERP analysis

The EEG was recorded from six cephalic Ag/AgCl electrodes F3, FZ, F4, C3, CZ, C4 referenced to an electrode placed on the nose. Fig. 2 shows electrode locations and their grouping into regions that were used in statistical analyses. The signal amplifier had a bandwidth of 0.3–100 Hz (DEYMED Diagnostic s.r.o., Czech Republic). The EEG was recorded at a 3000-Hz sampling rate.

The data were processed with Matlab release 2019b (Mathworks, USA). In the recorded EEG, the frequencies above 40.0 Hz were removed using a digital filter (implemented in EEGLab, Delorme and Makeig, 2004). Therefore, the spectral content of the analyzed EEG was 0.3–40.0 Hz. The EEG signal was downsampled to 300 Hz and epoched. The epoch started 100 ms before and ended 1000 ms after the vowel or tone onset; mean voltage of the prestimulus part (from − 100 ms to 0 ms) was subtracted from every epoch. The individual ERPs were calculated as an average of epochs with absolute amplitude under 90 μV. This procedure rejected about 25% of epochs; Table 2 shows the average number and the range of preserved epochs pooled across infants and channels. The level of signal to noise ratio for sensory ERP was determined by plus/minus procedure (Schimmel, 1967). We rejected 38 (out of 240) ERPs with SNR lower than one from further processing. The ERPs were additionally digitally filtered off-line by a low-pass Savitzky-Golay filter (Press et al., 1992, first polynomial order, window of 21 samples) to make responses better readable.

2.5. Statistical models

Data were analyzed with linear mixed-effects models using the packages lmer() and lmerTest() in R (Bates et al., 2015; Kuznetsova et al., 2017; R Core Team, 2016). One model was fitted for onset ERP, one for offset ERP, one for early MMR, and one for late MMR. The data entered in the model were ERP or MMR amplitudes averaged across trials per infant, dimension, electrode/scalp region, and stimulus type. The fixed and random-effects structures of each model are described in the respective Results subsections. In case of significant interactions, comparisons of the estimated 95% and 90% confidence intervals were done to localize the effect.

3. Results

3.1. ERPs: neural processing of stimulus physical properties

To test whether infants adequately processed the acoustic difference between the physically distinct stimuli, we compared the ERPs elicited by the acoustically different stimuli, i.e. averaging across all identical tokens with the exception of the first stimulus in each roving series. The ERPs were assessed in two 200-ms windows: an onset window 200–400 ms after vowel or tone onset, and an offset window 250–450 ms after vowel or tone offset. The window latencies were based on visual inspection of the grand-average waveforms, whereby the largest peak after stimulus onset was identified to lie at about 300 ms post-onset; and the largest peak after stimulus offset at about 350 ms after vowel or tone offset. The onset windows were aligned to vowel or tone onsets (i.e. the onset window in the speech stimulus was the onset of the V segment in the CV syllable) and were compared across stimuli that varied in their spectral properties. The offset windows were aligned to vowel and tone offsets and were as follows: 470–670 ms after stimulus onset for both the (medium-long) [a] and [ɛ] stimuli, 430–630 ms after onset for the short [ɛ] stimuli, and 610–810 ms after onset for the long [ɛː] stimuli. Offset responses were compared both across stimuli that varied in spectrum and across stimuli that varied in duration. The onset and offset responses were computed from ERP waveforms averaged across trials per infant, stimulus type, and electrode location, as areas under curve (AUC, in μV * ms) and submitted to the linear mixed models. The grand average ERPs are plotted in Fig. 3.

For the onset ERPs the model estimated the following parameters: the main effects of Domain (speech vs. nonspeech, coded as − 0.5 vs. +0.5) and Spectrum ([a] vs. [ɛ(ː)] including the short, intermediate and long variants of [ɛ], coded as − 0.5 vs. +0.5) and their interaction, the main effects of three location parameters, namely, Anteriority (central vs. frontal, coded as − 0.5 vs. +0.5), Laterality (with two contrasts, namely, left and right vs. midline, coded as − 0.25 and − 0.25 vs. +0.5, and left vs. right, coded as − 0.5 vs. +0.5), and their respective two- and three -way interactions with Domain and Spectrum. The model fitted per-participant random intercepts and random slopes for Spectrum. The offset model had the same predictors and random effects as the onset model, with main (fixed and random) and interaction effects of an additional parameter Duration (median-centred, coding 360 ms, 220 ms, and 180 ms, as +1, − 0.2, and − 0.6, respectively). The fixed-effects outputs are given in Table 3. In both models, the intercept was reliably larger than zero, indicating that overall, there was a meaningful, positive-going response after both stimulus onset and offset, averaging to AUC of 124 μV*ms and 58 μV*ms, respectively.

Both the onset and the offset response were affected by Domain: speech stimuli yielded larger onset and offset responses than nonspeech stimuli. Also, for both the onset and offset ERP, there were main effects of Laterality and Anteriority, but as these do not address any of our research questions we do not discuss them further.

More importantly for the present questions, there were significant three-way interactions involving Domain. For the onset response, Domain interacted with Spectrum and Anteriority. Table 4 lists the means and standard errors of the modelled means for each stimulus type in each condition for the onset and offset ERP; Fig. 4 depicts the means along with their confidence intervals. The left-hand graph in Fig. 4 shows that the [ɛ/ɛː] speech stimuli yielded larger response than the [a] speech stimuli (while no such differences were detected in nonspeech), in the central region. For the offset response, Domain interacted with Duration and Laterality. The right-hand graph in Fig. 4 shows that on the midline channels, longer speech stimuli yielded a larger offset response than shorter speech stimuli, while no such effect was seen in the nonspeech stimuli or on the lateral channels.

3.2. MMR: neural encoding of stimulus category

In order to test infants' mental encoding of sounds across domains we compared their neural responses to identical stimuli in different functional contexts. Difference waves were calculated by subtracting the ERP elicited by a stimulus when it served as a standard (namely, the last two tokens in a row of 4–8 identical stimuli) from the ERP elicited by the same physical stimulus when it served as a deviant (namely, the first token in the row). These difference waves allowed us to quantify abstract processing of the stimuli beyond their physical properties, i.e. to assess whether and to what extent a physically identical stimulus was processed specifically to the functional/sequential context in which it occurred (i.e. fulfilling the role of a standard versus a deviant). We computed the AUC of the difference wave in two time windows whose latencies were based on visual inspection of the grand-averaged data and are in line with the early and late MMR windows used in previous studies: an early MMR 80–220 ms after change onset, and a late MMR 500–700 ms after change onset. ‘Change onset’ corresponded to vowel and tone onset in the spectral domain, and to the short vowel and short tone offset in the durational domain. To increase the signal to noise ratio (which, compared to the primary ERPs became low due to a lower number of epochs averaged), we pooled central and frontal channels sharing laterality (i.e. F3 & C3, Fz & Cz, and F4 & C4).

Deviant identities were coded as follows. The spectral deviation from [fa] to [fɛ] (and alike for nonspeech stimuli) was coded as a change “to E” and the spectral deviation from [fɛ] to [fa] as a change “from E”; and alike for the nonspeech stimuli. Similar coding was adopted for deviant changes on the durational dimension, such that the durational deviation from [fɛː] to [fɛ] was coded as a change “to E”, and the durational deviation from [fɛ] to [fɛː] was coded as a change “from E”; and alike for the nonspeech stimuli. Fig. 5 plots the grand average difference waves.

Linear mixed effects models estimated the main effects of Domain (speech vs. nonspeech, coded as − 0.5 vs. +0.5), Dimension (duration vs. spectrum, coded as − 0.5 vs. +0.5), Deviant (to-E vs. from-E, coded as − 0.5 vs. +0.5), and all their two- and three-way interactions. The MMR models also included the main effect of Laterality (with two contrasts, namely, left and right vs. midline, coded as − 0.25 and − 0.25 vs. +0.5, and left vs. right, coded as − 0.5 vs. +0.5) and its respective two-, three-, and four-way interactions with Domain, Dimension, and Deviant. The models included per-participant random intercepts and random slopes for Dimension and Deviant, and their interaction.

Table 5 lists the output. For the early MMR, the first Laterality contrast turned out significant showing that the amplitude of the early MMR was smaller on the midline than laterally. For the late MMR, there was a significant three-way interaction of Domain, Dimension, and Laterality as well as a significant three-way interaction of Dimension, Deviant, and Laterality, both of which are licenced by a significant higher-order interaction.

The four-way interaction of Domain, Dimension, Deviant, and Laterality (left vs right) turned out significant for both the early and the late MMR. To unpack the interaction, we inspected the modelled means and compared them across the two Deviants in all conditions; Fig. 6 plots the means and 95% confidence intervals for the early MMR and Table 6 lists the means and standard errors for both the early and the late MMR. The pairwise comparisons show that in the speech domain the from-E, i.e. long, duration deviant yields a more negative MMR than the to-E, i.e. short, duration deviant on the left hemisphere (comparison of 95% c.i.s) and on the midline (comparison of 90% c.i.s). In the speech domain but this time on the spectral dimension, the to-E, i.e. [ɛ], spectral deviant yields a more negative MMR than the from-E, i.e. [a], spectral deviant (comparison of 90% c.i.s). Interestingly, the entire 95% c.i. of the [ɛ] spectral deviant on the left hemisphere is below zero, i.e. is reliably negative, arguably indexing a (relatively) mature MMR response – this is the only condition that elicits a mismatch negativity, i.e. MMN. For the late MMR, only the durational condition in speech shows a significant directional asymmetry in the left hemisphere (comparison of 90% c.i.s).

4. Discussion

4.1. Primary ERP responses

Hearing simple consonant-vowel syllables or inharmonic tone complexes elicited an automatic sensory response in newborns' brains. This means that newborn infants neurally process auditory stimuli both when they are speech and when they are nonspeech. Furthermore, the sensory responses elicited by the vowels were larger than those elicited by the complex tones. This indicates specialized cortical tuning to speech at the very level of its basic building blocks, and further extends the earlier documented infants' preferences for listening to larger chunks of speech versus analogue non-speech stimuli. Also, these automatic sensory responses elicited in sleep demonstrate a neural parallel to the earlier found behavioural preferences for speech over nonspeech in awake newborns' (Vouloumanos and Werker, 2007).

As evidenced by the triple interactions involving Domain for the onset and the offset ERP responses, the spectral difference between [a] and [ɛ] was reflected in significantly different onset ERP responses to [a] versus [ɛ] in the speech condition at central channels, and the durational difference between short [ɛ] and long [ɛː] was reflected in different offset ERPs to short versus long vowels at midline. This means that besides speech eliciting stronger neural responses than nonspeech in general, the acoustic differences between stimuli in terms of the first three formants, as well as in terms of duration, were more accurately processed when the stimuli were speech and less so when they were nonspeech. The more distinct acoustic response to the formant and duration differences in speech might be explained by a finer (experience based or innate) cortical tuning to speech. Alternatively, the more accurate processing in speech could be due to differential stimulus complexity across our stimulus sets. The speech stimuli were spectrally richer such that higher formants above F3 (which were not present in the non-speech condition) could have contributed to the perceived difference between [a] and [ɛ]. In a similar fashion, the fact that the vowels were preceded by a fricative consonant of constant duration might have facilitated the processing of the duration difference between short [ɛ] and long [ɛː] as compared to the short and long tones presented in isolation.

The topographical distribution of the auditory responses across the two domains, with midline and central regions reflecting robust processing of the acoustic content of linguistic stimuli, suggests a somewhat integrated processing pathway for speech. Thus, not only do speech stimuli differ from nonspeech analogues in that they are processed more robustly overall, but they also seem to activate other neural populations, whose specialisation remains to be determined.

4.2. Mismatch responses

The mismatch responses (MMR) patterned differently for speech than for nonspeech. The processing of speech sounds was asymmetrical: left-laterally, the [a] to [ɛ] change resulted in a more negative response than the [ɛ] to [a] change (and the [a] to [ɛ] change in speech was also the only condition that brought about a reliably negative MMN), and the [ɛ] to [ɛː] change resulted in a more negative response than the [ɛː] to [ɛ] change (and this durational asymmetry was observed also on the midline). Our first prediction that speech stimuli, unlike nonspeech, will yield a more mature MMR response is thus, partially, borne out. As the directional, left-lateralized asymmetries occurred both for the spectral and for the durational dimension in speech, our second prediction about vowel length having a developmental precedence over vowel quality is not supported.

The lateralization of the speech processing asymmetries to the left hemisphere adds to previous literature on hemispheric specialization for speech. Studies on the neural development of phoneme processing suggest that segmental speech processing starts bilaterally and only after the sixth month of an infant's development comes to be left-lateralized to resemble the hemispheric specialization found in adults (Arimitsu et al., 2011; Sato et al., 2012), although there are indications of left-hemisphere advantage in much younger infants (Dehaene-Lambertz and Baillet, 1998). Neurolinguistic studies with infants typically do not examine directionality effects in speech sound processing and therefore any subtle lateralization effects (corresponding to maturation) might have been previously obscured. Further work, with e.g. multichannel EEG that enables to more accurately localize sources of neural activity, is needed to confirm (or disprove) the lateralization of directional asymmetries detected here.

The newborns' left-lateralized asymmetries between the vowel quality deviants are reminiscent of the asymmetries previously reported for adults in some languages (e.g., Lahiri and Reetz, 2010, but see Mitterer, 2011, for counterevidence). Recall that in the present experiment, a change from [fa] to [fɛ] elicited a more robust negative mismatch response than a change from [fɛ] to [fa]. Although for instance German adults sometimes show similar directional effects for comparable vowel contrasts (e.g. Scharinger et al., 2012), Czech adults' neural discrimination of [fa]-[fɛ] exhibits an asymmetry in the opposite direction (Nudga et al., 2021). According to the Featurally Underspecified Lexicon (FUL, Eulitz and Lahiri, 2004) the specificity of speech sounds' mental representations determines whether and how much a sound is predictive, i.e. whether and how much its replacement by another speech sound violates a listener's expectation and causes an MMN. Assuming acquired, i.e. language-specific, phonological representations, Nudga et al., 2021 argued that Czech /a/ is phonologically underspecified (for backness), causing that a change from the un(der)specified, less predictive /a/ to a fully-specified /ɛ/ does not violate an expectation in Czech adult listeners while a reverse change does. The Czech newborns in the present study had an MMN asymmetry in the opposite direction, which indicates that their processing – quite expectedly – was not affected by the phonological makeup of the Czech vowel system.

Although lacking phonological knowledge, newborns do have some prior experience with speech in terms of its acoustics. An account that addresses asymmetries shaped by phonetic biases in young infants has been proposed by Polka and Bohn (2003, 2011). These authors' Natural Referent Vowel framework refers to vowels' articulatory-acoustic properties and argues that peripheral vowels such as [a], [i], and [u], thanks to their unique articulatory-acoustic characteristics, are stable points in the vowel space and universally serve as perceptual anchors. Other authors (Schwartz et al., 2005) argued that it is the acoustic properties of peripheral vowels, namely the closeness of neighbouring vowel formants, i.e., focalization, which makes vowels like [a], [i], and [u] perceptually prominent. According to the NRV (Polka and Bohn, 2011), a young infant who has been exposed to spoken language will discriminate a change from a nonperipheral [ɛ] to a peripheral [a] more robustly than a change in the reverse direction (this directionality effect has been confirmed in the meta-analysis by Tsuji and Cristia, 2017), while later in development these auditorily-conditioned asymmetries may leave way for language-specific patterns (Pons et al., 2012; but see Tsuji and Cristia, 2017, who did not find an interaction effect of age and nativeness). The asymmetry detected in the present experiment with newborns is not in line with the asymmetry predicted by the NRV.

We propose that the perceptual asymmetry in our newborn data might be caused by differential learning stages for each of the two vowel categories. The concentrated energy at about 1 kHz – which is a frequency band that still has a relatively good chance of propagating into the womb (Richards et al., 1992) – makes [a] perceptually more salient (and especially so in utero) than [ɛ] whose energy is dispersed across a wider frequency range. Furthermore, in spoken Czech tokens of /a/ are more frequent than tokens of /ɛ/ (by about 15–20%, ORAL v1, 2019). Hypothetically, fetuses who had been exposed to somewhat vaguely audible and slightly less frequent [ɛ]s and to better audible and more frequent [a]s, could have more readily started to form a perceptual category for /a/ than for /ɛ/. Upon hearing tokens of [a] after birth (in the present experiment), the neonate listeners recognized a previously encountered, and perhaps somewhat 'primitively' learned /a/-category, and could establish a memory trace for it during the experimental paradigm such that with every upcoming trial they anticipated hearing that vowel category (in line with the predictive coding theory, see Winkler and Czigler, 2012). When the [a] stimulus changed into [ɛ], their memory trace of /a/ was violated, as reflected in a strong MMN response to the [ɛ] deviant. On the contrary, upon hearing tokens of [ɛ], there was no category to be recognized, no memory trace was built up during a repeated presentation of [ɛ]s, such that a change from [ɛ] to [a] did not violate any expectation. This is why the [a] deviant resulted in a much weaker MMR than the [ɛ] deviant.

As a reviewer pointed out, phonetic warping-induced asymmetries are addressed by the Native Language Magnet model (NLM, Kuhl, 1991; Kuhl et al., 2008). According to the NLM, the internal structure of segmental speech categories (acquired through exposure), which comprises a best instance of the category - the prototype, and its variants, predicts directional asymmetries. The prototype acts as a perceptual magnet: when the prototype is heard first, the difference between it and a non-prototypical variant is perceived as smaller than when the variant is heard first. Even though the present experiment tested discrimination across two adult categories, one could potentially argue that the fetuses/newborns would warp the entire vowel space of [a]s and [ɛ]s into a single 'protocategory' (as also modelled by Chládková et al., 2020). Assuming such a protocategory in which the focal and more frequent [a] is more prototypical than the less salient and less frequent [ɛ], the NLM would predict better discrimination for a change from [ɛ] to [a] than vice versa, which is the opposite of what we found in the newborns' MMR. At this point, it is unclear whether the newborns perceived [a] and [ɛ] as instances of one protocategory, or as two different – and perhaps differently well-warped – adult categories, or whether they were still blank-slates without any prior warping/categorization having occurred.

Although neither of the two influential models of early speech perception, the NRV and the NLM, did specifically refer to prenatal development, it is intriguing that the asymmetries we detected here with newborns run counter to both the phonetically-based NRV's as well as the categorization-based NLM's predictions. Potentially, the language-general biases predicted by the NRV (Polka and Bohn, 2011), or the prototype-driven biases predicted by the NLM (Kuhl et al., 2008), might occur in slightly older infants after sufficient experience with speech ex utero, or, they might, after all, be language- or phoneme-specific (i.e. not applicable to infants acquiring Czech, or to [mid-]low vowels such as [ɛ] and [a]).

Could the present reversal of NRV- or NLM-predicted asymmetries be attributed to having measured discrimination at the neural level? The NRV was proposed to explain asymmetries found in infants' behavioral discrimination (Polka and Bohn, 2011), and the NLM's predictions for asymmetries were, too, mostly attested with behavioral methods (e.g. Moon et al., 2013; but note that Kuhl et al., 2008, explicitly propose that exposure to native language will result in language-specific processing at the neural level). Neural discrimination patterns are typically – at least to some extent – reflected in behavioral measures of vowel discrimination (see the review in Näätänen, 2001, for early work and e.g. Virtala et al., 2018, 2020 for more recent work). If anything, neural change detection precedes behavioral change detection: Tremblay et al. (1998) showed that after training the MMN to phoneme contrasts improved even though such improvement was not detectable at the level of behavior. As for the case of perceptual asymmetries, one may expect that neurally a contrast could yield a similarly strong MMN in both directions of change, yet behaviorally one direction would be discriminated more readily than the other direction (see Polka et al., 2021, who did not detect a MMN asymmetry for [y]-[u] in adults who typically have an asymmetry in behavioral tasks). A complete reversal of an asymmetry across the neural MMR and behavioral level would mean that a direction of change that is poorly detectable by a neural, pre-attentive, index of discrimination is well detectable behaviorally, which we consider rather unlikely. We thus like to argue that the dissonance between ours and NRV- or NLM-like asymmetries is not due to the use of MMR in the present experiment. Nevertheless, it is still worth exploring further whether measures of neural speech processing other than the MMR reveal (other kinds of) perceptual asymmetries: a potential measure to look at is the oscillatory theta or gamma activity. In infants theta activity seems to reflect general phonetic decoding of speech irrespective of comprehension, and gamma activity relates to processing of language-specific/syllabic information (Ortiz Barajas et al., 2021): in potential future work on infants' perceptual asymmetries and neural oscillations one might hypothesize that phonetically-shaped biases be reflected in the theta band (which is also what Polka et al., 2021, observed in adults) and categorically-shaped biases in the gamma-band.

Let us now turn to the perceptual asymmetry in the durational vowel contrast. As far as contrasts such as /a/-/ɛ/ are concerned, the literature relatively widely documents and theorizes about the asymmetries. Much less is known about potential asymmetries in the perception of length. Previous studies, mostly with adults, typically (though not always) find that listeners more robustly process changes from short to long stimuli than vice versa, probably because an addition of information is more readily detectable than a loss of it (Jaramillo et al., 1999; Ylinen et al., 2006). The short-to-long easy detectability does not, however, explain the perceptual patterns of the newborns in the present study. There was a more negative MMR to a short-to-long deviant than to a long-to-short deviant in speech, but no such effect was seen in the non-speech stimuli which differed in duration in exactly the same way as the speech sounds. Therefore, the asymmetry in speech might not be (entirely) due to the immediate stimulus acoustic properties.

Although the NRV (Polka and Bohn, 2011) addresses vowel length only briefly, it suggests that short vowels may – similarly to focal vowels – serve as perceptual anchors, such that discriminating a change from a long to a short vowel would then be easier than vice versa. Regarding the prototype-biases postulated by the NLM (Kuhl et al., 2008), the more frequent short vowel could be considered more prototypical than the long one, thus predicting better discrimination from long to short than vice versa. The durational asymmetry that we found here is, again, a reversal of the asymmetry postulated by the NRV and the NLM frameworks.

As in the case of the spectral contrast, the MMR asymmetry for vowel length could possibly reflect the newborns' prior experience with durationally varying speech input and differential degree of warping for the short versus the long categories. In Czech, short vowels are more frequent than long vowels (ORAL v1, 2019). Also, considering absolute duration scales, it appears that tokens of Czech short vowels are rather compactly clustered around a prototypical short value, while tokens of Czech long vowels are a bit more widely spread around a particular long duration value (Lehiste, 1970; Paillereau and Chládková, 2019), and this differential dispersion in the short and the long category might in prenatal IDS be even larger than in ADS (Chládková et al., 2019)[4]. A developing fetus who encounters many similarly short vowels and fewer variously long vowels might more readily recover and start warping the narrowly-defined underlying short category as opposed to a broadly-distributed underlying long category. In the current experiment, upon hearing tokens of the (partially) warped short category, the newborn listener might establish a memory trace and build up a prediction, which – when violated by a long stimulus – results in a more negative MMR response than does a reverse violation.

In this section we speculated about the possible cause of MMR asymmetries in vowel perception at birth. We argued that the newborns' speech-specific left-lateralized asymmetries in neural discrimination of vowels may reflect a more advanced stage of perceptual warping for some vowel categories than for others. At this point however, we cannot rule out an alternative explanation that the perceptual patterns seen here are universal, innate, and have no relation to the language spoken in the babies' environment. In that respect, the asymmetries could simply reflect infants' general preference for, or tuning into, speech over nonspeech. To what extent prenatal experience with speech leads to early perceptual categorization of the ambient speech sounds remains to be tested.

4.3. Limitations and future research

A potential methodological confound to the speech vs. nonspeech sensory ERP comparison is the extent to which the material in each domain was informative. We aimed to present the same acoustic patterns in the context of speech and nonspeech stimuli. Since vowels hardly ever occur as isolated segments in natural speech, we used the smallest typically occurring speech units – consonant-vowel monosyllables. Besides strengthening the 'speech-likeliness' of the stimuli, the syllable-initial fricative might have, however, provided supporting acoustic information. The [f] had invariant duration and frication formant, which could have served as reference points for perceptual discrimination and categorization of the immediately following vowel. Potentially, the initial fricative might have contributed to the stronger primary ERPs to acoustic stimulus differences in speech as compared to nonspeech. (However, it is less likely that the fricative contributed to the asymmetries in MMR within the speech condition, since all speech stimuli began with an identical fricative.) To resolve whether the stronger primary responses to speech were domain-specific, or were driven by the extra acoustic information, a possible follow-up experiment could employ nonspeech stimuli that entail referencing information, roughly comparable to an initial consonant in CV syllables.

We proposed that prenatal experience with listening to speech could have resulted in the asymmetries observed in this study. To assess the plausibility of prenatal vowel learning, one needs to test infants, and/or near-term fetuses with different language backgrounds. However, those populations are particularly demanding to recruit and test (and especially so for a cross-language design), and have noisier data than older listeners. To that end, computational modelling may provide valuable insights, leading to informed hypotheses for experiments with such young humans. Seebach et al. (1994) tested whether the English plosive place of articulation is learnable prenatally. A neural network, modelling the fetal hearing capacities and intrauterine sound properties, was exposed to realizations of English /pa/-/ta/-/ka/. The network came to differentiate the three-way categorical contrast and even generalized the acquired knowledge to untrained /ba/-/da/-/ga/. One could thus hypothesize (and test) that near-term

---

[4] However, when logarithmic scales for duration are considered, which are perhaps more psychoacoustically plausible than absolute scales (Abel, 1972), the dispersion in short and long vowel categories seems comparable.

fetuses, exposed to English would perceptually discriminate (and perhaps even categorize) the three-way consonantal place distinction.

Using two-layer neural networks, research in our lab showed that Spanish but not Czech near-term virtual fetuses will form two separate 'protocategories' for [a]- and [ɛ]-like vowels (Chládková et al, 2020). In a cross-linguistic experiment, Spanish-exposed newborns would thus be predicted to discriminate [a] and [ɛ] more robustly than Czech-exposed newborns. Considering the present MMR asymmetries in Czech newborns' processing of [fa]-[fɛ], a more robust discrimination by Spanish newborns might mean an overall more negative and/or symmetric MMR. Supposedly, fetuses and newborns exposed to Spanish, which, unlike Czech, does not contrast short and long vowels, might have an attenuated MMR to a vowel length distinction, such as the [fɛ]-[fɛː] used here. Alternatively, one could test a single language group of newborns on changes within- and across adult categories: Czech newborns could be tested on their neural discrimination of variants of /ɛ/ and variants of /a/. If prenatal phonetic warping takes place – perhaps for /a/ if it is focalization, or perhaps for /ɛ/ if it is the lowest formant that matters in utero – one could expect to find prototype-like directional effects in the strength of MMR for that particular vowel category (Kuhl et al., 2008). A cross-sectional study comparing newborns to older infants (e.g. 6- and 12-month olds) could help identify the degree of warping/categorization at birth (if any).

The present study does not answer the question of whether segmental speech sound learning starts already in utero: the hypothesized, input saliency-based, difference in newborns' categorization of phonemic vowel length versus phonemic vowel quality was not found. However, the unexpected left-lateralized directional asymmetry of the newborns' neural discrimination for both phoneme contrasts offers new insights into the earliest stages of speech learning: it has lead us to speculate about a potential scenario of prenatal speech development which is testable in future work. Ultimately, experiments that compare newborn infants or fetuses from different language environments are crucial in order to answer questions about the effects that prenatal experience has on the formation of speech sound categories in the young infant.

5. Conclusions

We pursued the question of whether humans might learn about the speech sounds of their language before they are born, and whether some speech categories are learned earlier than others. Sleeping newborns listened to native-language speech sound differences, namely, [fɛ]-[fa] and [fɛ]-[fɛː], and to similar nonspeech stimuli, namely, inharmonic complex tones.

Sensory ERPs to the speech stimuli were overall stronger and more reliably reflected the differences in stimulus spectral and durational characteristics than did the ERPs to nonspeech. The mismatch responses differed across domains, indicating left-lateralized directional asymmetries in the processing of speech stimuli. Contrary to our predictions, we did not detect any differences between the two types of speech contrasts, suggesting that phonemic vowel length and vowel spectral quality contrasts were, by the third day of life processed comparably.

The most intriguing result were the directional asymmetries in speech. Left-laterally, infants had a more mature mismatch response to a change from [fa] to [fɛ] than vice versa, and to a change from [fɛ] to [fɛː] than vice versa. We proposed a hypothetical scenario of how prior experience could modulate newborn speech sound processing, arguing that the newborns' perceptual asymmetries reflected differential degrees of prenatal perceptual warping of /a/

versus /ɛ/, and of /ɛ/ versus /ɛː/. To what extent our proposal is realistic – and in general, whether naturalistic speech sound category formation occurs before birth – remains to be addressed in future work.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data Availability

The data that support the findings of this study are available at OSF https://osf.io/b9txc/.

Acknowledgments

This study was funded by Charles University grant Primus/17/HUM/ 19. KC was further supported by the Czech Science Foundation grant 18- 01799S, and JU and JK by Charles University project Progres Q40/7. We are grateful to Zuzana Ocelakov ′ a ′ (Charles University) for help with experimental setup, to Gabor Stefanics (University of Zurich) for sharing hands-on experience with newborn ERP research, to the staff of the maternity ward at the University Hospital in Hradec Kralové for cooperation, and to all the participating families. We are grateful to two anonymous reviewers for their thorough comments on earlier versions of the article.

References

- Abboub, N., Nazzi, T., Gervain, J., 2016. Prosodic grouping at birth. Brain Lang. 162, 46–59. Abel, S.M., 1972. Duration discrimination of noise and tone bursts. J. Acoust. Soc. Am. 51 (1972), 1219–1223. https://doi.org/10.1121/1.1912963.
- Arimitsu, T., Uchida-Ota, M., Yagihashi, T., Kojima, S., Watanabe, S., Hokuto, I., Ikeda, K., Takahashi, T., Minagawa-Kawai, Y., 2011. Functional hemispheric specialization in processing phonemic and prosodic auditory changes in neonates. Front. Psychol. 2, 202. https://doi.org/10.3389/fpsyg.2011.00202.
- Bates, D., Mächler, M., Bolker, B., Walker, S., 2015. Fitting linear mixed-effects models using lme4. J. Stat. Softw. 67 (1), 1–48. https://doi.org/10.18637/jss.v067.i01. Boersma, P., Weenink, D. (1992–2020). Praat: Doing Phonetics by Computer. Available at: 〈http://www.praat.org〉 (accessed 22-Nov-2018).
- Cheour, M., Ceponiene, R., Lehtokoski, A., Luuk, A., Allik, J., Alho, K., Näätänen, R., 1998. Development of language-specific phoneme representations in the infant brain. Nat. Neurosci. 1, 351–353.
- Cheour, M., Martynova, O., Näätänen, R., Erkkola, R., Sillanpää M., Kero, P., Raz, A., Kaipio, M.L., Hiltunen, J., Aaltonen, O., Savela, J., Hämäläinen, H., 2002. Psychobiology: speech sounds learned by sleeping newborns. Nature 415 (6872), 599–600.
- Chládková, K., Nudga, N., Boersma (2020). A model of prenatal vowel learning. In: Proceedings of the Cognitive Science Society annual meeting, pp. 599–604.
- Chládková, K., Černá, M., Paillereau, N., Skarnitzl, R., Oceláková, Z. (2019). Prenatal infant-directed speech: vowels and voice quality. In: Proceedings of ICPhS 2019, 1525–1529.

- DeCasper, A.J., Lecanuet, J.-P., Busnel, M.-C., Granier-Deferre, C., Maugeais, R., 1994. Fetal reactions to recurrent maternal speech. Infant Behav. Dev. 17, 159–164.
- Dehaene-Lambertz, G., Baillet, S., 1998. A phonological representation in the infant brain. Neuroreport 9, 1885–1888.
- Delorme, A., Makeig, S., 2004. EEGLAB: an open source toolbox for analysis of singletrial EEG dynamics including independent component analysis. J. Neurosci. Methods 134 (1), 9–21.
- Eulitz, C., Lahiri, A., 2004. Neurobiological evidence for abstract phonological representations in the mental lexicon during speech recognition. J. Cogn. Neurosci. 16, 577–583.
- Goudbeek, M., Swingley, D., Smits, R., 2009. Supervised and unsupervised learning of multidimensional acoustic categories. J. Exp. Psychol.: Hum. Percept. Perform. 35, 1913–1933.
- Granier-Deferre, C., Ribeiro, A., Jacquet, A., Bassereau, S., 2011. Near-term fetuses process temporal features of speech. Dev. Sci. 14 (2), 336–352.
- Graven, S.N., Browne, J.V., 2008. Auditory development in the fetus and infant. Newborn Infant Nurs. Rev. 8 (4), 187–193.
- Haenschel, C., Vernon, D.J., Dwivedi, P., Gruzelier, J.H., Baldeweg, T., 2005. Eventrelated brain potential correlates of human auditory sensory memory-trace formation. J. Neurosci. 25 (45), 10494–10501.
- Jaramillo, M., Alku, P., Paavilainen, P., 1999. An event-related potential (ERP) study of duration changes in speech and non-speech sounds. NeuroReport 10, 3301–3366.
- Kisilevsky, B.S., Hains, S.M., Brown, C.A., Lee, C.T., Cowperthwaite, B., Stutzman, S.S., Huang, H., 2009. Fetal sensitivity to properties of maternal speech and language. Infant Behav. Dev. 32 (1), 59–71.
- Kremláček, J., Kreegipuu, K., Tales, A., Astikainen, P., Poldver, N., Näätänen, R., Stefanics, G., 2016. Visual mismatch negativity (vMMN): a review and meta-analysis of studies in psychiatric and neurological disorders. Cortex 80, 76–112. https://doi.org/10.1016/j.cortex.2016.03.017.
- Kuhl, P.K., 1991. Human adults and human infants show a “perceptual magnet effect” for the prototypes of speech categories, monkeys do not. Percept. Psychophys. 50 (2), 93–107.
- Kuhl, P.K., Conboy, B.T., Coffey-Corina, S., Padden, D., Rivera-Gaxiola, M., Nelson, T., 2008. Phonetic learning as a pathway to language: new data and native language magnet theory expanded (NLM-e). Philos. Trans. R. Soc. B 363, 979–1000. https://doi.org/10.1098/rstb.2007.2154.
- Kuznetsova, A., Brockhoff, P.B., Christensen, R.H.B., 2017. LmerTest package: tests in linear mixed effects models. J. Stat. Softw. 82 (13), 1–26. https://doi.org/10.18637/jss.v082.i13.
- Lahiri, A., Reetz, H., 2010. Distinctive features: phonological underspecification in processing. J. Phonetics 38, 44–59.
- Lecanuet, J.-P., Carolyn Granier-Deferre, Decasper, A., R Maugeais, JA Andrieu, Busnel, C. (1987). Fetal Perception and Discrimination of Speech Stimuli; Demonstration by Cardiac Reactivity; Preliminary Results. 305(5), 161–164.
- Lehiste, I., 1970. Suprasegmentals. MIT Press, Cambridge, MA. Lüdecke, D., 2018. ggeffects: tidy data frames of marginal effects from regression models. J. Open Source Softw. 3 (26), 772. https://doi.org/10.21105/joss.00772.
- Maddieson, I., 1986. The size and structure of phonological inventories: analysis of UPSID. Exp. Phonol. 105–123.

- Mahmoudzadeh, M., Dehaene-Lambertz, G., Fournier, M., Kongolo, G., Goudjil, S., Dubois, J., … Wallois, F. (2013). Syllabic discrimination in premature human infants prior to complete formation of cortical layers. Proceedings of the National Academy of Sciences, 110(12), 4846–4851.
- Martynova, O., Kirjavainen, J., Cheour, M., 2003. Mismatch negativity and late discriminative negativity in sleeping human newborns. Neurosci. Lett. 340 (2), 75–78. https://doi.org/10.1016/s0304-3940(02)01401-5.
- Maurer, U., Bucher, K., Brem, S., Brandeis, D., 2003. Development of the automatic mismatch response: from frontal positivity in kindergarten children to the mismatch negativity. Clin. Neurophysiol. 114, 808–817.
- May, L., Gervain, J., Carreiras, M., Werker, J.F., 2018. The specificity of the neural response to speech at birth. Dev. Sci. 21 (3), e12564.
- Mehler, J., Jusczyk, P., Lambertz, G., Halsted, N., Bertoncini, J., Amiel-Tison, C., 1988. A precursor of language acquisition in young infants. Cognition 29 (2), 143–178.
- Mitterer, H., 2011. The mental lexicon is fully specified: Evidence from eye-tracking. J. Exp. Psychol.: Hum. Percept. Perform. 37, 496–513.
- Monahan, P.J., Idsardi, W.J., 2010. Auditory sensitivity to formant ratios: toward an account of vowel normalization. Lang. Cogn. Process. 25 (6), 808–839. https://doi.org/10.1080/01690965.2010.490047.
- Moon, C., Cooper, R.P., Fifer, W.P., 1993. Two-day-olds prefer their native language. Infant Behav. Dev. 16, 495–500.
- Moon, C., Lagercrantz, H., Kuhl, P.K., 2013. Language experienced in utero affects vowel perception after birth: a two-country study. Acta Paediatr. 102 (2), 156–160.
- Mueller, J.L., Friederici, A.G., Männel, C., 2012. Auditory perception at the root of language leanring. PNAS 109, 15953–15958.
- Näätänen, R., 2001. The perception of speech sounds by the human brain as reflected by the mismatch negativity (MMN) and its magnetic equivalent (MMNm). Psychophysiology 38 (1), 1–21. https://doi.org/10.1111/1469-8986.3810001.
- Nenonen, S., Shestakova, A., Huotilainen, M., Näätänen, R., 2005. Speech-sound duration processing in a second language is specific to phonetic categories. Brain Lang. 92, 26–32. ORAL v1: Czech National Corpus – ORAL v1 (2019). Institute of the Czech national corpus, Faculy of Arts, Charles University, Prague. url: 〈https://kontext.korpus.cz/〉 [accessed 23–12-2019].
- Nudga, N., Urbanec, J., Oceláková, Z., Kremláček, J., Chládková, K., 2021. Neural processing of spectral and durational changes in speech and non-speech stimuli: an MMN study with Czech adults. Front. Hum. Neurosci. 15, 643655 https://doi.org/10.3389/fnhum.2021.643655.
- Ortiz Barajas, M.C., Guevara, R., Gervain, J., 2021. The origins and development of speech envelope tracking during the first months of life. Dev. Cogn. Neurosci. 48, 100915 https://doi.org/10.1016/j.dcn.2021.100915.
- Paillereau, N., Chládková, K., 2019. Spectral and temporal characteristics of Czech vowels in spontaneous speech. AUC Philol. 2019, 77–95. https://doi.org/10.14712/24646830.2019.19.
- Partanen, E., Kujala, T., Näätänen, R., Liitola, A., Sambeth, A., Huotilainen, M., 2013. Learning-induced neural plasticity of speech processing before birth. Proc. Natl. Acad. Sci. U.S.A. 110, 15145–15150.
- Peter, V., Kalashnikova, M., Santos, A., Burnham, D., 2016. Mature neural responses to infant-directed speech but not adult-directed speech in pre-verbal infants. Sci. Rep. 6, 34273. https://doi.org/10.1038/srep34273.

- Polka, L., Bohn, O.-S., 2003. Asymmetries in vowel perception. Speech Commun. 41 (1), 221–231.
- Polka, L., Bohn, O., 2011. Natural referent vowel (NRV) framework: an emerging view of early phonetic development. J. Phonetics 39, 467–478.
- Polka, L., Molnar, M., Zhao, C., Masapollo, M., 2021. Neurophysiological correlates of asymmetries in vowel perception: an English-French cross-linguistic ERP study. Front. Hum. Neurosci. 15, 607148 https://doi.org/10.3389/fnhum.2021.607148.
- Pons, F., Albareda-Castellot, B., Sebastian-Gall ′ ′es, N., 2012. The interplay between input and initial biases: Asymmetries in vowel perception during the first year of life. Child Dev. 83 (3), 965–976.
- Press, W.H., Teukolsky, S.A., Vetterling, W.T., Flannery, B.P., 1992. Savitzky-Golay smoothing filters. Numerical Recipes in C: The Art of Scientific Computing. Cambridge University Press, Cambridge, pp. 650–655.
- Querleu, D., Renard, X., Versyp, F., Paris-Delrue, L., Cr`epin, G., 1988. Fetal hearing. Eur. J. Obstetr. Gynecol. Reprod. Biol. 28 (3), 191–212.
- R Core Team (2016). R: a language and environment for statistical computing. R Foundation for Statistical Computing, Vienna. [〈www.r-project.org〉].
- Richards, D.S., Frentzen, B., Gerhardt, K.J., McCann, M.E., Abrams, R.M., 1992. Sound levels in the human uterus. Obstetr. Gynecol. 80, 186–190.
- Sato, H., Hirabayashi, Y., Tsubokura, H., Kanai, M., Ashida, T., Konishi, I., Maki, A., 2012. Cerebral hemodynamics in newborn infants exposed to speech sounds: a whole-head optical topography study. Hum. Brain Mapp. 33 (9), 2092–2103.
- Scharinger, M., Herrmann, B., Nierhaus, T., Obleser, J., 2014. Simultaneous EEG-fMRI brain signatures of auditory cue utilization. Front. Neurosci. 8, 137.
- Scharinger, M., Idsardi, W.J., Poe, S., 2011. A comprehensive three-dimensional cortical map of vowel space. J. Cogn. Neurosci. 23, 3972–3982.
- Scharinger, M., Monahan, P.J., Idsardi, W.J., 2012. Asymmetries in the processing of vowel height. J. Speech Lang. Hearing Res. 55, 903–918.
- Schimmel, H., 1967. The (+-) reference: accuracy of estimated mean components in average response studies. Science 157, 92–94.
- Schwartz, J.-L., Abry, C., Boë, L.-J., Valleé, N., 2005. The dispersion-focalization theory of sound systems. J. Acoust. Soc. Am. 117 (4), 2422. https://doi.org/10.1121/1.4786487.
- Seebach, B.S., Intrator, N., Lieberman, P., Cooper, L.N., 1994. A model of prenatal acquisition of speech parameters, 91, 7473–7476. https://doi.org/10.1073/pnas.91.16.7473.
- Shahidullah, S., Hepper, P.G., 1994. Frequency discrimination by the fetus. Early Hum. Dev. 36 (1), 13–26.
- Thiede, A., Virtala, P., Ala-Kurikka, I., Partanen, E., Huotilainen, M., Mikkola, K., Kujala, T., 2019. An extensive pattern of atypical neural speech-sound discrimination in newborns at risk of dyslexia. Clin. Neurophysiol. 130 (5), 634–646.
- Tremblay, K., Kraus, N., McGee, T., 1998. The time-course of auditory perceptual learning: Which comes first, the chicken or the egg. NeuroReport 9, 3557–3560.
- Tsuji, S., Cristia, A. (2017). Which acoustic and phonological factors shape infants vowel discrimination? Exploiting natural variation in InPhonDB. In: Proceedings of Interspeech 2017, pp. 2108–2112. doi: 〈10.21437/Interspeech.2017–1468〉 .
- Virtala, P., Partanen, E., Tervaniemi, M., Kujala, T., 2018. Neural discrimination of speech sound changes in a variable context occurs irrespective of attention and explicit

awareness. Biol. Psychol. 132, 217–227. https://doi.org/10.1016/j. biopsycho.2018.01.002.

- Virtala, P., Talola, S., Partanen, E., Kujala, T., 2020. Poor neural and perceptual phoneme discrimination during acoustic variation in dyslexia. Sci. Rep. 10, 8646. https://doi. org/10.1038/s41598-020-65490-3.
- Vouloumanos, A., Werker, J.F., 2007. Listening to language at birth: evidence for a bias for speech in neonates. Dev. Sci. 10, 159–164.
- Wanrooij, K., Boersma, P., Van Zuijen, T., 2014. Fast phonetic learning occurs already in 2-to-3-month old infants: An ERP study. Front. Psychol. 5, 77.
- Weikum, W.M., Oberlander, T.F., Hensch, T.K., Werker, J.F., 2012. Prenatal exposure to antidepressants and depressed maternal mood alter trajectory of infant speech perception. Proc. Natl. Acad. Sci. U.S.A. 109, 17221–17227.
- Winkler, I., Czigler, I., 2012. Evidence from auditory and visual event-related potential (ERP) studies of deviance detection (MMN and vMMN) linking predictive coding theories and perceptual object representations. Int. J. Psychophysiol. 83, 132–143.
- Ylinen, S., Shestakova, A., Huotilainen, M., Alku, P., Na¨¨ at¨ anen, R., 2006. Mismatch negativity (MMN) elicited by changes in phoneme length: a crosslinguistic study. Brain Res. 1072, 175–185.
- Zhao, T., Moon, C., Lagercrantz, H., Kuhl, P.K., 2011. Prenatal motherese? Newborn speech perception may be enhanced by having a young sibling. Psi Chi J. Undergrad. Res. 16, 90–94. K

Table 1. Infant demographics per the between-subject condition, domain.

condition	*n* included (*n* tested)	*n* per sex	age at experiment: mean (range)	birth weight: mean (range)
speech	54 (60)	30 F, 24 M	57 h (30 – 108)	3395 g (2720–4420)
nonspeech	50 (60)	25 F, 25 M	54 h (28 – 87)	3363 g (2620–4100)

Table 2. Average count, minimum, and maximum of preserved epochs, pooled across infants and channels, for each stimulus type in the ERP and MMR analyses.

			ERPs			MMR: Deviants			MMR: Standards		
Domain	**Presentation block**	**Stimulus**	**min**	**mean**	**max**	**min**	**mean**	**max**	**min**	**mean**	**max**
speech	spectrum	ɛ	188	388	505	38	77	101	81	155	201
		a	181	382	499	35	77	100	67	153	200
	duration	ɛ	173	396	502	30	79	100	79	157	199
		ɛː	189	390	489	32	78	98	69	156	198
nonspeech	spectrum	ɛ	180	393	495	38	78	99	68	156	198
		a	208	387	495	36	78	99	82	155	199
	duration	ɛ	185	380	492	38	76	100	72	151	198
		ɛː	201	375	486	31	75	99	77	150	196

Table 3. Fixed-effects output of the linear mixed models for the onset and offset ERP. Bold font marks effects with p below 0.05.

Parameter	Onset ERP					Offset ERP				
	Estimate	SE	df	t	p	Estimate	SE	df	t	p
Intercept	**123.876**	**11.711**	**109.157**	**10.578**	**<0.001**	**57.683**	**13.459**	**107.939**	**4.286**	**<0.001**
Domain (-speech, +nonspeech)	**-89.014**	**23.422**	**109.157**	**-3.800**	**<0.001**	**-58.24**	**26.918**	**107.939**	**-2.164**	**0.033**
Spectrum (-a, +e)	1.683	20.524	104.609	0.082	0.935	17.471	21.090	103.510	0.828	0.409
LateralityA (-lateral, +midline)	10.351	14.406	2204.536	0.718	0.473	**60.892**	**15.597**	**2104.588**	**3.904**	**<0.001**
LateralityB (-left, + right)	**25.425**	**12.476**	**2204.536**	**2.038**	**0.042**	1.907	13.508	2104.588	0.141	0.888
Anteriority (-central, +frontal)	**-30.073**	**10.187**	**2204.536**	**-2.952**	**0.003**	**59.474**	**11.029**	**2104.588**	**5.393**	**<0.001**
Domain * Spectrum	-45.120	41.048	104.609	-1.099	0.274	64.027	42.180	103.510	1.518	0.132
Domain * LateralityA	-54.618	28.812	2204.536	-1.896	0.058	-56.877	31.195	2104.588	-1.823	0.068
Domain * LateralityB	-9.511	24.952	2204.536	-0.381	0.703	36.330	27.016	2104.588	1.345	0.179
Domain * Anteriority	-13.940	20.373	2204.536	-0.684	0.494	-8.420	22.058	2104.588	-0.382	0.703
Spectrum * LateralityA	14.609	28.812	2204.536	0.507	0.612	14.957	31.632	2104.588	0.473	0.636
Spectrum * LateralityB	3.683	24.952	2204.536	0.148	0.883	5.826	27.394	2104.588	0.213	0.832
Spectrum * Anteriority	-30.501	20.373	2204.536	-1.497	0.135	10.144	22.367	2104.588	0.454	0.650
Domain * Spectrum * LateralityA	-29.405	57.625	2204.536	-0.510	0.610	44.325	63.264	2104.588	0.701	0.484
Domain * Spectrum * LateralityB	-6.750	49.904	2204.536	-0.135	0.892	-27.457	54.788	2104.588	-0.501	0.616
Domain * Spectrum * Anteriority	**101.350**	**40.747**	**2204.536**	**2.487**	**0.013**	10.955	44.734	2104.588	0.245	0.807
Duration						4.780	14.877	100.361	0.321	0.749
Domain * Duration						-1.775	29.754	100.361	-0.060	0.953
Duration * LateralityA						22.193	23.016	2104.588	0.964	0.335
Duration * LateralityB						-8.201	19.932	2104.588	-0.411	0.681

Table 4. Modelled means and standard errors (SE) for onset ERP in the central and frontal region, and for offset ERP in the left, midline, and right region. Significance of pairwise comparisons (p.c.) across Stimulus types is indicated by asterisks: ** marks mutually exclusive means in the 95% confidence intervals estimated for each deviant type, * marks mutually exclusive means in 90% confidence intervals. Calculation of confidence intervals: 95% c.i. = mean ± 1.96SE, 90% c.i. = mean ± 1.645SE. The means and SEs were estimated using the ggeffects R package (Lüdecke, 2018, function ggpredict).

	Onset ERP							Offset ERP									
Region →		**central**			**frontal**				**left**			**midline**			**right**		
Domain	**Stimulus**	**mean**	**SE**	**p.c.**	**mean**	**SE**	**p.c.**	**Stimulus**	**mean**	**SE**	**p.c.**	**mean**	**SE**	**p.c.**	**mean**	**SE**	**p.c.**
speech	a	128.443	11.711	**	145.9	15.5		short	45.7	15.2		75.2	24.9	**(lo.-sh.)	39.8	22.9	
	ɛ	182.417	21.319		118.7	27.7		medium	50.7	13.4		98.3	20.7	*(lo.-me.)	31.4	19.2	
								long	65.6	21.3		167.7	36		6.1	33	
nonspeech	a	107.3	26.4		60.2	35		short	-44.4	34.3		-9.8	56.3		-30.6	51.7	
	ɛ	76.2	48.2		49.3	62.7		medium	-38.4	30.3		-11.9	46.8		-17.8	43.3	
								long	-20.4	48.3		-18.3	81.7		20.7	74.8	

Table 5. Fixed-effects output of the linear mixed models for the early and the late MMR. Bold font marks effects with p below 0.05.

Parameter	Early MMR					Late MMR				
	Estimate	SE	df	t	p	Estimate	SE	df	t	p
Intercept	26.916	16.448	90.652	1.636	0.105	8.168	32.397	100.224	0.252	0.802
Domain (-speech +nonspeech)	28.050	32.896	90.652	0.853	0.396	23.630	64.794	100.224	0.365	0.716
Dimension (-duration +spectrum)	-13.848	31.791	90.342	-0.436	0.664	8.855	71.592	105.082	0.124	0.902
Deviant (-toE +fromE)	18.690	32.537	106.125	0.574	0.567	29.245	54.834	101.250	0.533	0.595
LateralityA (-lateral +midline)	**-37.640**	**17.794**	**2020.00**	**-2.115**	**0.035**	-24.677	34.790	2020.000	-0.709	0.478
LateralityB (-left +right)	28.404	15.410	2020.00	1.843	0.065	1.532	30.129	2020.000	0.051	0.960
Domain * Dimension	-24.203	63.581	90.342	-0.381	0.704	23.321	143.184	105.082	0.163	0.871
Domain * Deviant	1.860	65.075	106.125	0.029	0.977	65.264	109.669	101.250	0.595	0.553
Dimension * Deviant	88.837	69.387	105.271	1.280	0.203	65.800	113.763	102.553	0.578	0.564
Domain * LateralityA	-13.235	35.587	2020.00	-0.372	0.710	62.340	69.580	2020.000	0.896	0.370
Domain * LateralityB	-28.206	30.819	2020.00	-0.915	0.360	37.686	60.258	2020.000	0.625	0.532
Dimension * LateralityA	-18.029	35.587	2020.00	-0.507	0.613	48.439	69.580	2020.000	0.696	0.486
Dimension * LateralityB	-34.191	30.819	2020.00	-1.109	0.267	-76.024	60.258	2020.000	-1.262	0.207
Deviant * LateralityA	-65.521	35.587	2020.00	-1.841	0.066	-65.846	69.580	2020.000	-0.946	0.344
Deviant * LateralityB	25.909	30.819	2020.00	0.841	0.401	0.883	60.258	2020.000	0.015	0.988
Domain * Dimension * Deviant	-148.284	138.775	105.271	-1.069	0.288	-128.812	227.526	102.553	-0.566	0.573
Domain * Dimension * LateralityA	27.083	71.175	2020.00	0.381	0.704	29.797	139.160	2020.000	0.214	0.831
Domain * Dimension * LateralityB	-21.371	61.639	2020.00	-0.347	0.729	**-248.684**	**120.516**	**2020.000**	**-2.063**	**0.039**
Domain * Deviant * LateralityA	52.623	71.175	2020.00	0.739	0.460	192.463	139.160	2020.000	1.383	0.167
Domain * Deviant * LateralitvB	7.167	61.639	2020.00	0.116	0.907	-73.560	120.516	2020.000	-0.610	0.542

Table 6. Modelled means and standard errors (SE) for early and late MMR at left, midline, and right channels. Significance of pairwise comparisons (p.c.) across Deviants is indicated by asterisks: ** marks mutually exclusive means in the 95% confidence intervals estimated for each deviant type, * marks mutually exclusive means in 90% confidence intervals. Calculation of confidence intervals: 95% c.i. = mean ± 1.96SE, 90% c.i. = mean ± 1.645SE. The means and SEs were estimated using the ggeffects R package (Lüdecke, 2018, function ggpredict).

				Early MMR									Late MMR								
Laterality →				left			midline			right			left			midline			right		
Domain	Dimension	Deviant		mean	SE	p.c.	mean	SE	p.c.	mean	SE	p.c.	mean	SE	p.c.	mean	SE	p.c.	mean	SE	p.c.
speech	duration	to-E	[short]	34.9	16.4	**	55.3	24.2	*	46.8	22.5		124.4	32.4	**	-12.4	47.5		-16.1	44.2	
		from-E	[long]	-54.0	36.2		-43.0	53.8		42.7	50.0		-51.2	67.0		-64.7	102.7		6.4	95.0	
	spectrum	to-E	[ɛ]	-65.6	32.4	*	-32.3	51.2		-14.8	47.2		-104.8	69.6		33.6	104.2		-36.6	96.7	
		from-E	[a]	82.3	77.7		9.6	117.8		93.0	109.2		107.4	153.4		-82.8	231.4		52.9	214.6	
non-speech	duration	to-E	[short]	25.8	37.2		35.6	54.8		82.0	51.0		-86.7	73.4		-74.5	107.6		98.6	100.2	
		from-E	[long]	66.3	82.1		30.3	121.8		83.6	113.2		-1.9	151.7		68.8	232.4		54.0	215.0	
	spectrum	to-E	[ɛ]	49.3	73.0		2.0	115.4		-7.9	106.4		30.8	157.2		43.9	235.0		-77.8	218.2	
		from-E	[a]	38.1	175.3		7.4	265.8		78.9	246.3		90.4	346.0		54.7	521.8		39.4	483.9	

Figure 1. Oscillograms and spectrograms of the speech (I.) and nonspeech stimuli (II.). The depicted amplitude scale is relative, both the speech and nonspeech stimuli were presented at 67 dB SPL (as measured by a dummy head using infant earcouplers with fitted earplugs).

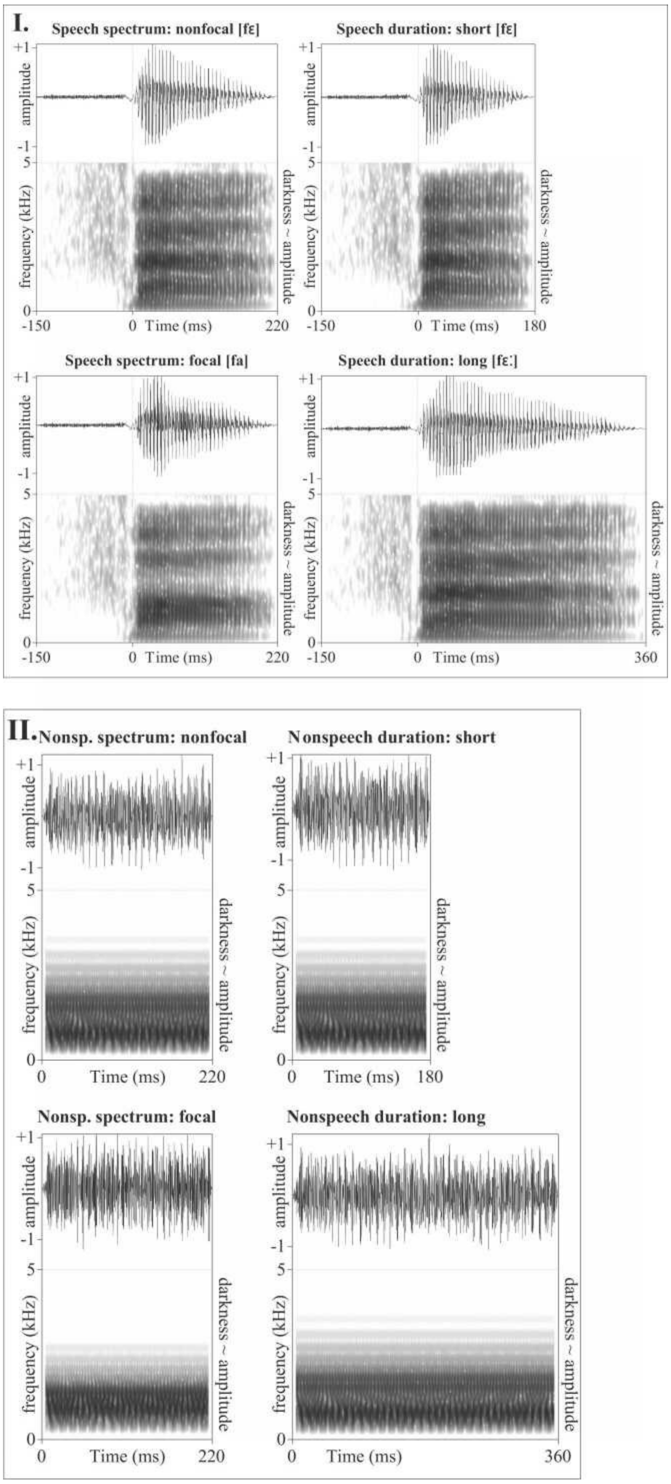


Figure 2. The recording sites and grouping of channels into 5 regions.

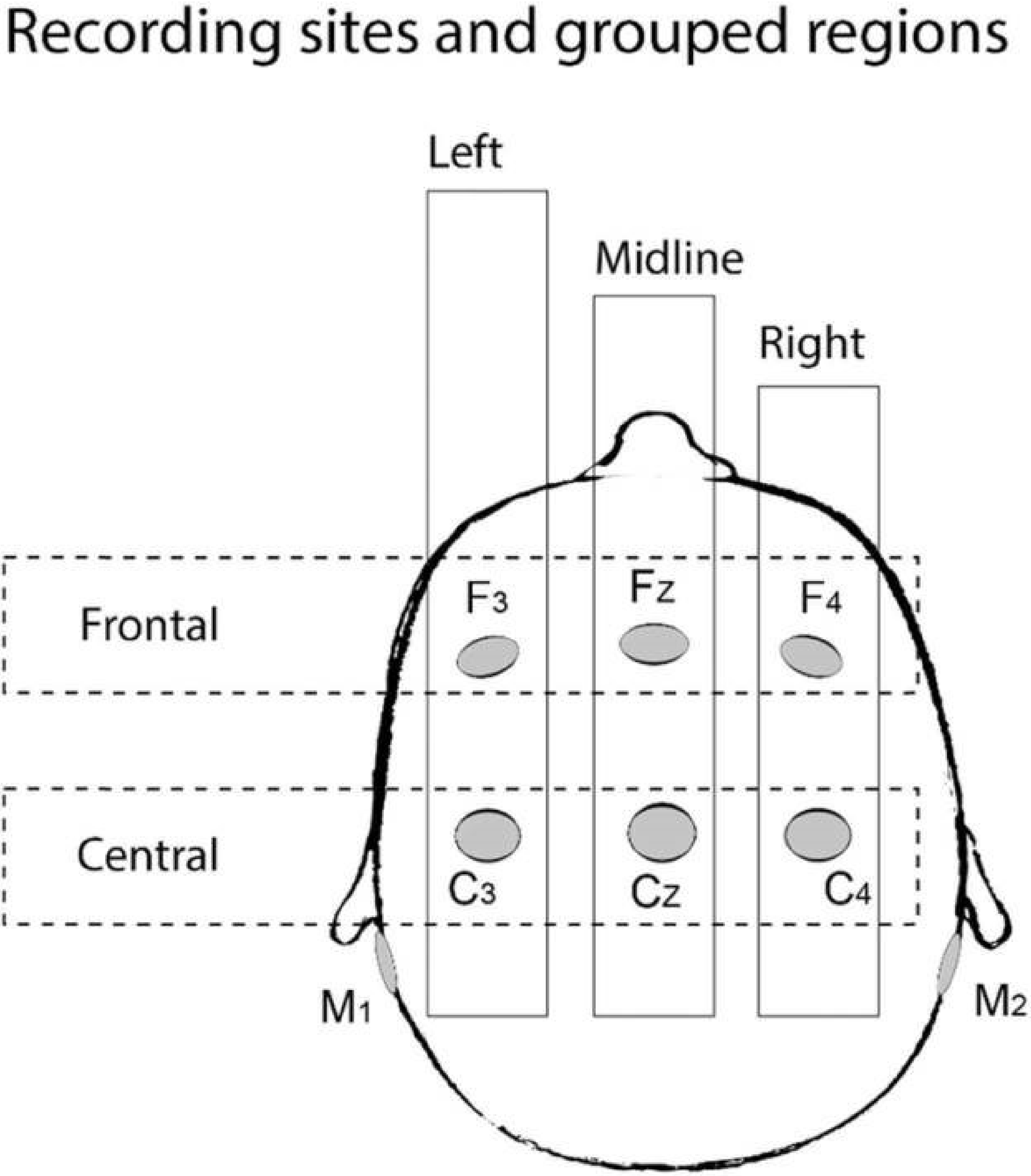


Figure 3. Upper five graphs: grand average ERPs to [a](-like) and [ɛ]/[ɛː](-like) stimuli. Lower five graphs: grand average ERPs to long, medium, and short stimuli. Shaded rectangles mark the analysis windows.

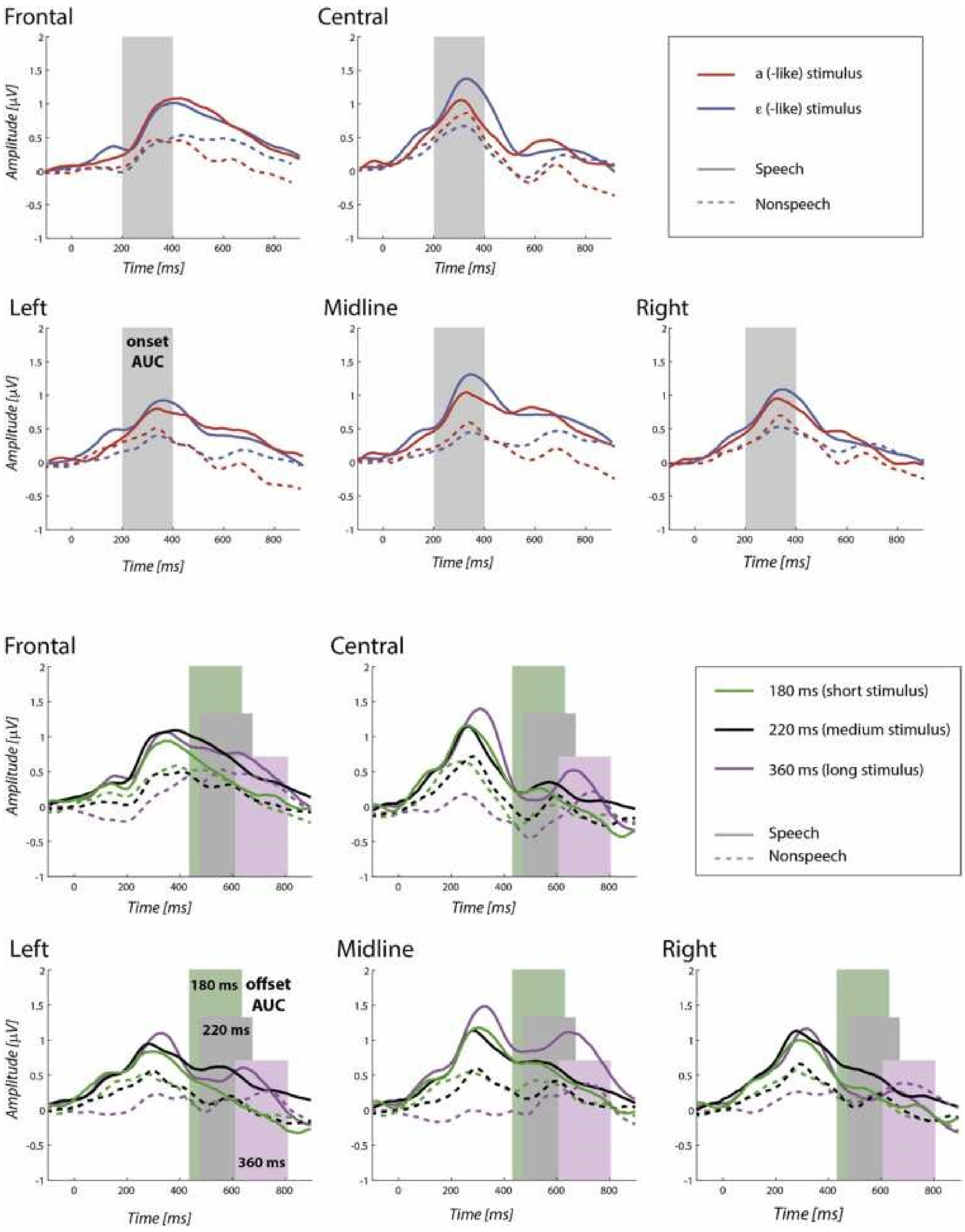


Figure 4. Modelled means and 95% CIs for the onset ERP (left) depicting the interaction of Domain, Spectrum, and Anteriority, and for the offset ERP (right) depicting the interaction of Domain, Duration, and Laterality. Colour coding aligns with the colours of the grand average ERP waves plotted in Fig. 3.

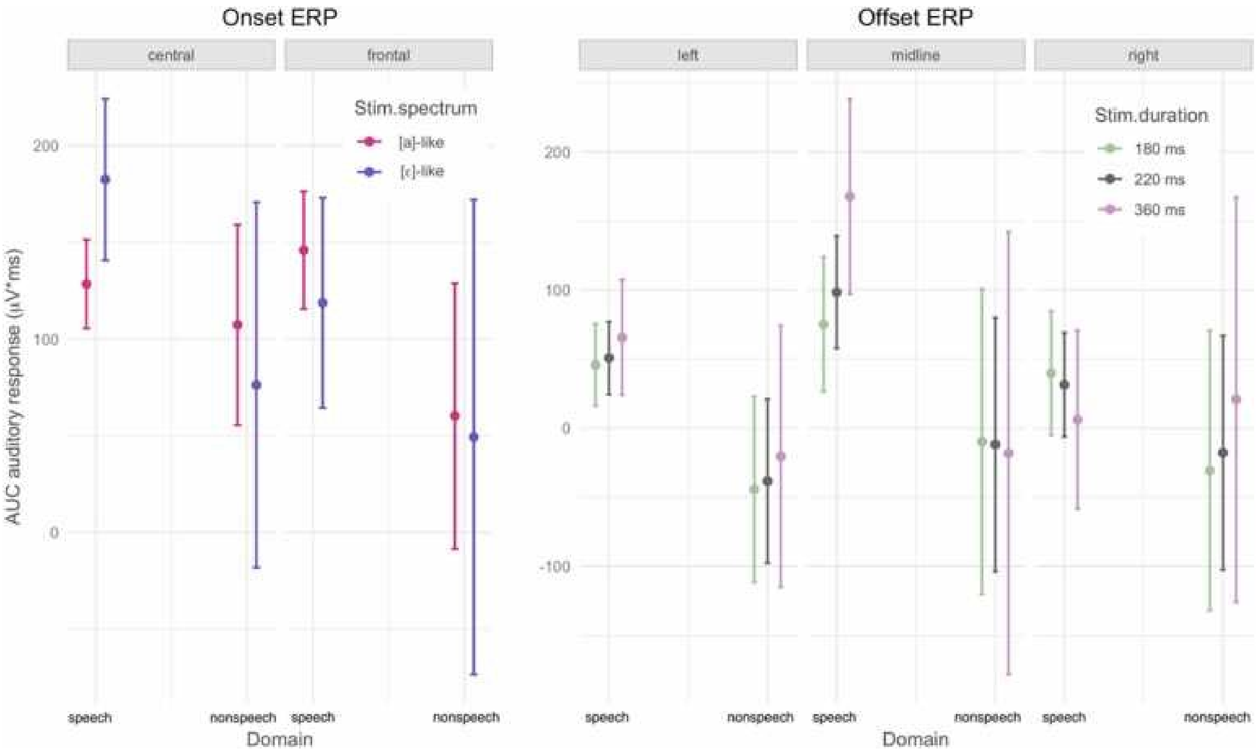


Figure 5. Grand average difference waves in the three scalp regions (for region visualization, see Fig. 2). Shading shows the early and late MMR analysis windows. Numbers in the top right corners show over how many participants averaging was done in each condition. The difference waves were computed from physically identical stimuli, e.g. the difference wave for the spectral “from-E” deviant was computed as: ERP for [a] as deviant minus ERP for [a] as standard, and the difference wave for the spectral “to-E” deviant was computed as: ERP for [ɛ] as deviant minus ERP for [ɛ] as standard, and likewise for the durational deviations between [ɛ] and [ɛː].

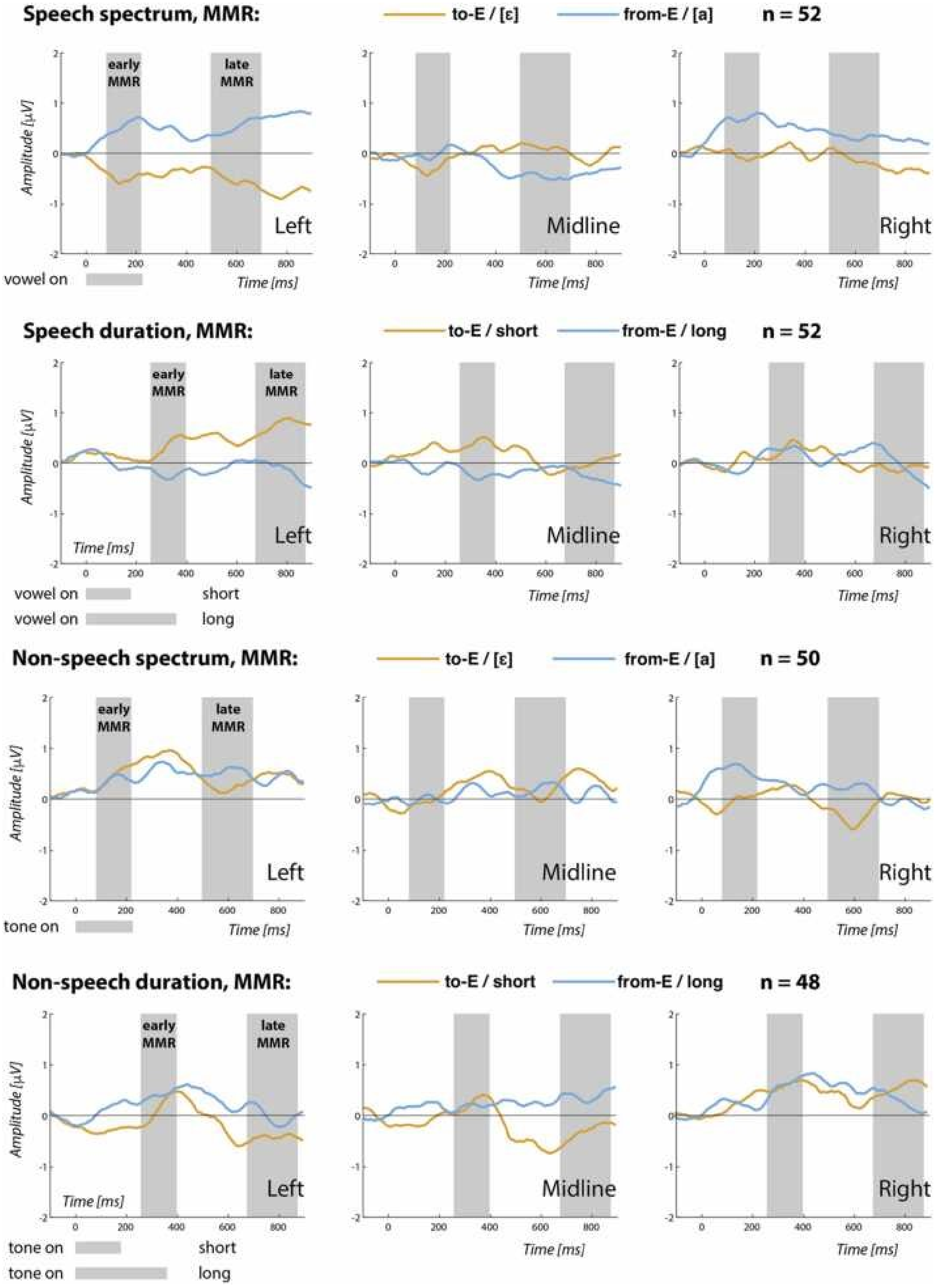
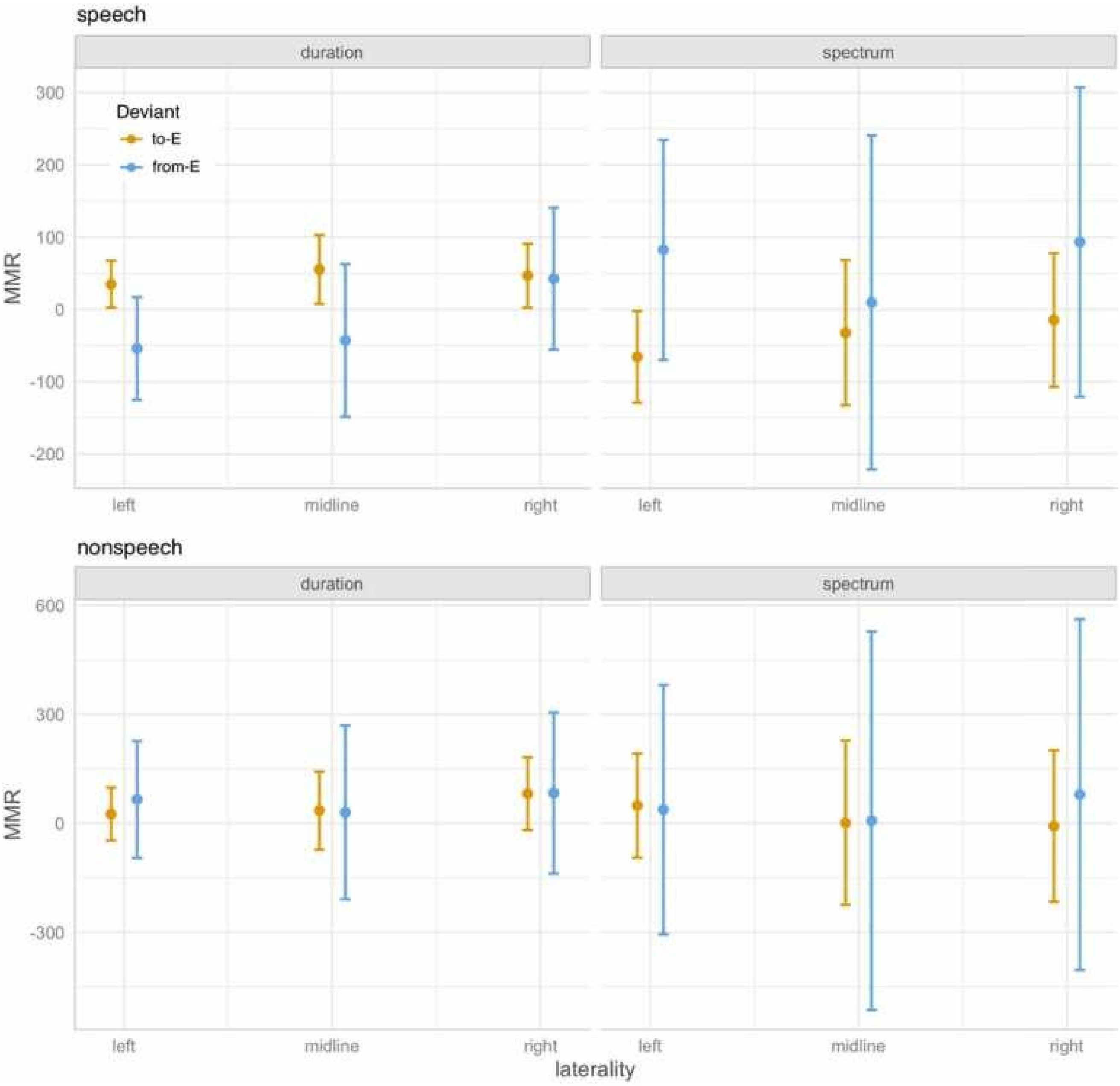


Figure 6. The modelled means and 95% c.i.s for the early MMR, unpacking the Domain * Dimension * Deviant * Laterality interaction. Deviant label "to-E" corresponds to the [ɛ] deviant on the spectral dimension and to the short deviant on the duration dimension, and deviant label "from-E" corresponds to the [a] deviant on the spectral dimension and to the long deviant on the duration dimension.



5.3 Neural processing of speech sounds at premature and term birth: ERPs and MMR between 32 and 42 weeks of gestation

Josef Urbanec[1], Kateřina Chládková[2,3], Jan Kremláček[1]

[1] Department of Medical Biophysics, Faculty of Medicine in Hradec Králové, Charles University, Czechia

[2] Institute of Czech Language and Theory of Communication, Faculty of Arts, Charles University, Prague, Czechia

[3] Institute of Psychology, Czech Academy of Sciences, Prague, Czechia

Corresponding author:

Kateřina Chládková

Institute of Czech Language and Theory of Communication

Faculty of Arts, Charles University

Nám. Jana Palacha 1-2

116 38, Prague Czechia

katerina.chladkova@ff.cuni.cz

Abstract

Prenatal listening experience reportedly modulates how humans process speech at birth, but little is known about how speech perception develops throughout the perinatal period. The present experiment assessed the neural event-related potentials (ERP) and mismatch responses (MMR) to native vowels in 99 neonates born between 32 and 42 weeks of gestation. The vowels elicited reliable ERPs in newborns whose conception age at time of experiment was at least 36 weeks and 1 day (36+1). The ERPs reflected spectral distinctions between vowel onsets from conception age 36+6 and durational distinctions at vowel offsets from conception age 37+6. Starting at age 40+4, there was evidence of neural discrimination of vowel length, indexed by a negative MMR response. The present findings extend our understanding of the earliest stages of speech perception development in that they pinpoint the ages at which the cortex reliably responds to the phonetic characteristics of individual speech sounds and discriminates a native phoneme contrast. The age at which the brain reliably differentiates vowel onsets coincides with what is considered term age in many countries (37+0 GA). Future studies should investigate to what extent the perinatal maturation of the cortical responses to speech sounds is modulated by the ambient language.

Keywords

newborn speech perception; event-related potentials; mismatch response; perinatal development; premature birth; vowels

1 Introduction

The attunement to the native language begins during the last period of intrauterine development, sometime after the 28th week of gestation when the auditory pathways are in place. Near-term fetuses and newborn infants recognize their mother's voice, the global characteristics of the language their mother spoke during pregnancy, as well as rhymes she recited during the last weeks of pregnancy (DeCasper & Fifer 1980, May et al 2018, DeCasper et al. 1994). The prenatal learning of spoken language is more intricate than pure remembering of global language patterns: there is evidence that by the time they are born humans have already started to generalise over the linguistic structures such as the intonational and rhythmic patterns specific to their native language (Mampe et al. 2009, Abboub et al. 2016). Moreover, studies indicate that the prenatal learning of native-language patterns might pertain even to smaller-sized structures such as the identities of individual vowels and syllables (Moon et al. 2013, Partanen et al. 2013, Chládková et al. 2021). While current behavioural and neuroimaging literature demonstrates that humans do have the various language-specific abilities at the time of birth, it still remains unknown when exactly during prenatal development the attunement to native linguistic patterns sets on. The aim of our experiment is to advance the current understanding of the very beginnings of spoken language development by testing at what gestational age the newborn cortex distinguishes between minimally contrastive native-language syllables.

The literature indicates that the ability to discriminate syllables develops sometime between the 28th and 35th week of gestation: fetuses stimulated with syllables [ba] and [bi] show behavioural signs of discriminating such stimuli at the 35th but not at the 28th week of gestational age (Lecanuet et al. 1987). Using an indirect measure of neural activity, the brain's hemodynamic response, Mahmoudzadeh et al. (2013) tested discrimination of consonant-vowel syllables in twelve preterm infants' born between 28 and 32 weeks of gestation age. Different patterns of hemodynamic activity were detected for stimulation with strings of repeating identical syllables [ga] compared to strings of oddball blocks with two different syllable identities [ga] and [ba]. Using data from the same experimental session, Mahmoudzadeh et al. (2017) measured the ERPs and reported neural discrimination of the (predictably occurring) changes in syllable identity (as well as speaker voice). The results suggested that the cortex of preterm newborns distinguishes between the two different syllables. Daneshvarfard et al. (2019) assessed the frequency following response (FFR) in the cortical auditory responses to strings of [ba] and [ga] in 16 preterm newborns born between 29 and 34 weeks of gestation. They found that the accuracy and the phase coherence of the response correlates with age, suggesting development of the frequency-following response across the tested preterm age range.

The period between approximately the 30th and 36th week of gestation reportedly marks a change in the cortical as well as subcortical processing of sounds. Starr et al. (1977) examined the auditory brainstem responses (ABRs) to nonspeech stimuli (clicks) in 42 newborn infants ranging in age between 25 and 44 weeks of gestation, and found that the ABRs stabilise in gestation week 36. It is in the same period when also the cortical responses, the auditory event-related potentials (ERPs) of prematurely-born neonates change in their appearance and come to resemble those of full-term newborns. This change in the ERPs is characterised by a shift from a dominant negative peak to a dominant positive peak at the latency of about 200–250 ms after the onset of an auditory (non-speech) stimulus (Roetteveel et al. 1987, Eggermont & Moore 2012). The auditory event-related potentials change significantly from birth up until adolescence. The auditory ERP waveforms in infants born extremely preterm at 24 weeks display a negative peak at about 200 ms post stimulus onset and a positive peak at about 600 ms, whose latency decreases with development. At term, it is the positive peak that comes to dominate the auditory cortical response with a latency of about 250 ms post stimulus onset while the negative component seen in extremely premature infants is no longer visible in the ERP waveform (Eggermont & Moore 2011). This dominance of the large positive peak at about 200 ms latency remains a characteristic of infant and toddler auditory ERP for at least several years; the negative N1 component, characteristic of auditory ERP in adults, fully develops only at about 5 to 6 years of age or even later (Lippé et al 2009, Ruhnau et al. 2011). The maturation of auditory ERPs is observed earlier at midline regions and later also at temporal sites (Guzzetta et al 2011). The degree to which ERPs are mature is affected by how stimuli are presented: an adult-like N1 can be observed at younger ages with longer inter-stimulus intervals and at older ages with shorter inter-stimulus intervals (Ruhnau et al. 2011).

While there are a number of studies that assessed the cortical processing of speech stimuli between preterm and fullterm newborns, they do not allow to make inferences about the developmental trajectory of cortical speech sound processing because comparison were made between fullterm infants and preterm infants at term age (Peña et al. 2012, François et al. 2019, Kostilainen et al. 2020). In order to pinpoint the age at which discrimination of native speech sounds starts to be reliably indexed by the auditory event-related brain potentials, our experiment assesses the event-related potentials in 99 newborns spanning gestation ages 32 to 42 weeks.

As to stimulus characteristics, prior research shows that newborns' brains process speech and nonspeech stimuli differently (when presented with continuous speech, May et al. 2018, but also when presented with isolated syllables, Chládková et al. 2021), one can thus expect that the developmental trajectory of auditory ERPs will differ between speech and nonspeech stimuli. Here we focus on the development of cortical processing specific to speech, which is modulated not only by auditory and neural maturation but also by prenatal speech input, and which may very likely differ from the development of cortical processing of non-speech. The present experiment aims to show when in gestational development the cortex discriminates between minimally distinct syllables of the ambient language.

The maturational stages of auditory ERPs have been relatively well documented for changes *between* infancy, toddlerhood, childhood, and adolescence as well as within adulthood (Wunderlich et al. 2006, Ruhnau et al. 2011, Mahajan & McArthur 2012, Tomé et al. 2015) but are considerably less well documented *within* infancy or *within* gestational development as such (Kushnerenko et al. 2002). Given that auditory ERPs to speech at birth have been repeatedly shown to correlate with later language outcomes and language-related disorders both in full-term and in premature infants (Thiede et al. 2019, Maitre et al 2013), it is necessary to have a more detailed understanding of how the cortical auditory processing develops in the earliest stages of development, and particularly so for speech sounds.

While it may take several years for the maturation for the primary auditory ERP components such as the N1 and P2 to complete, studies with young infants often focused on a secondary ERP measure, the mismatch response (MMR), as an index of auditory development, and speech perception development in particular. The MMR is assessed in a difference waveform obtained by subtracting the ERP to one type of stimulus (a frequently presented one) from an ERP to another type of stimulus (an infrequently presented one). While some consider

MMR an ontogenically early ERP response (Stefanics et al. 2007) others underline its status as being an investigator-developed construct as it is never measured directly from the scalp (unlike the N1 or P2 components) but only obtained through subtraction of the recorded ERPs (Eggermont & Moore 2011). Despite that, studies on auditory and speech processing with young infants or even fetuses largely rely on the MMR. Considering the MMR as an index of maturation might not be straightforward: it turns out that to reliably identify which factors affect the MMR polarity and latency in infants is not trivial, and at the same time, it becomes clear that age alone is not the primary modulating factor (Govaart, Dvořáková et al. 2023). Interpreting the MMR with reference to the primary ERP responses, thus allows to more comprehensively assess the development of early cortical processing of speech. To investigate how the developing cortex responds to different native speech sounds, we thus measure the primary auditory ERPs. To investigate phonetic discrimination beyond the primary sensory processing of acoustic stimulus differences, we measure the neural discrimination index, the MMR.

We presented sleeping newborns with trains of isolated vowels from their native language, Czech, which differed in spectral quality or in duration. As acoustic signals pass through the maternal tissue, abdomen, and bones, their spectral properties from about ~700 Hz and above are attenuated while durational properties are transmitted veridically (Richards et al. 1992, Granier-Deferre et al. 2011). One can assume that if the ambient language systematically differentiates vowels not only in terms of spectral properties but also in terms of duration by having short and long vowel categories (as Czech does), the developing fetus may more robustly sensitise to speech sound contrasts cued by duration. We thus predicted that Czech-exposed newborns may begin to differentiate differences in vowel duration earlier than vowel spectral properties, which might be indexed by a more mature MMR response and/or differences in the primary ERP responses. However, considering that vowel duration is cued at stimulus offset, and vowel spectral quality at stimulus onset, a confound comes to play whereby offset ERPs are reported to be in general weaker than onset ERPs (in adults, Baltzell & Billings 2014). To this end, the MMR will provide valuable insights into the neural discrimination of durational versus spectral vowel contrasts as it is not dependent solely on stimulus physical properties (unlike the ERPs) but also on the auditory system abstracting away from the immediate stimulus, building up predictions on the upcoming vowel identity and evaluating violations to those predictions (Garrido et al. 2009). Besides allowing us to trace the early development of cortical responses to speech sounds, the present experiment will enable us to compare the developmental trajectory across different types of speech stimuli.

The present study assesses sensory cortical processing of different native vowels as well as the neural index of phonetic discrimination. Tracing the brain's speech sound processing across neonate infants born between 32 and 42 weeks of gestation age will allow us to identify the ages at which the cortex reliably distinguishes across native speech sounds, and compare the maturation of speech-elicited ERPs to prior findings on perinatal auditory processing of nonspeech signals.

2 Method

2.1 Participants

A total of 102 infants were tested, 3 of them were excluded due to administration of unusual neonatal drugs, congenital malformation of the brain and cardiopulmonal resuscitation after delivery. Data of 99 infants were retained for analysis. Figure 1 shows their gestational age at birth and at time of experiment (termed conception age), sex, and the condition to which they were (randomly) assigned. The infants were born between the 32th and 42th gestation weeks and tested on the 3th day after birth (range 1 to 16 days, in the most preterm babies usually in the second week of life because of previous life support). Their birth weight ranged from 1500 to 4370 grams. All infants had 10-minute Apgar scores 8 or higher and passed the neonatal hearing test (typically administered the 3rd day after birth in term newborns). Newborns delivered by vacuum extraction or forceps were not recruited. The infants were born to women whose native language was Czech. The experiment was approved by the ethics committee of Havlíčkův Brod hospital, Czechia. Infants took part in the experiment following a parental written informed consent.

2.2 Stimuli and paradigm

Infants were assigned to one of two conditions, receiving either durational-change or spectral-change stimulation. The durational condition tested the contrast between [ɛ] and [ɛː] and the spectral change condition tested the contrast between [ɛ] and [a], both vowel contrasts representing a phonemic change in Czech, the infants' native language. The vowels were from natural recordings of a Czech female speaker who produced a series of [f]-vowel monosyllables. For each vowel category, the most clear and prototypically-sounding vowel was extracted as the middle 50% portion of the vocalic interval and edited for duration using PSOLA in Praat (Boersma & Weenink 1992–2024). The first three formant values of [ɛ] were 755 Hz, 1646 Hz, and 2710 Hz, and the first three formant values of [a] were 864 Hz, 1287 Hz, and 2831 Hz. The duration of the short [ɛ] and [a] was 180 ms, and the duration of the long [ɛː] was 360 ms. The stimuli were presented at 65 dB SPL via insert earphones attached on the inner side of infant ear couplers. Figure 2 illustrates the setup.

Each infant listened to two oddball blocks in which the standard and the deviant swapped roles. That is, for the spectral condition, one of the blocks had [ɛ] as standards and [a] as deviants, and vice versa for the other block, with the order of blocks being counterbalanced. The durational change blocks were analogous, one block with [ɛ] as standard and [ɛː] as deviant, the other with the role of the two vowels reversed. Each block contained a total of 843 stimuli out of which 120 were deviants (deviant probability being 14.2%). A block always started with 9 standards and subsequently there were 3 to 9 standards between successive deviants. The stimulus onset asynchrony jittered randomly between 990 ms and 1190 ms (in 10-ms steps). Each block lasted 15.3 minutes. There was a brief break between the blocks to allow switching stimulation and checking electrode impedances with the infant kept asleep.

2.3 EEG recording & procedure

EEG was recorded from 6 scalp electrodes placed at the locations F3, Fz, F4, C3, Cz, and C4 according to the international 10/20 system. External electrodes were placed on the nose (online reference), on the face (grounding electrode), and on the chest or a hand to monitor ECG. EEG was recorded at a 1000-Hz sampling frequency. Impedances were kept below 50 kΩ. Infants were tested while asleep; infant state was monitored by a video camera. The experimenter (the first author) and in most cases also the infant's mother were silently present in the testing room during the whole recording session. Figure 2 shows the recording setup in one of the infants.

2.4 EEG preprocessing

The signal amplifier's bandwidth spanned from 0.3 to 100 Hz (DEYMED Diagnostic s.r.o., Czech Republic). Data processing was carried out using Matlab release 2023a (Mathworks, USA). Frequencies exceeding 40.0 Hz in the recorded EEG were eliminated using a digital filter (using the inverse Fast Fourier Transformation, implemented in EEGLab as eegfiltfft, Delorme and Makeig 2004). As a result, the spectral composition of the analyzed EEG was constrained to 0.3–40.0 Hz. The EEG signal underwent epoching, commencing 100 ms before and concluding 1000 ms after the vowel onset. The average voltage of the prestimulus segment (from −100 ms to 0 ms) was subtracted from each epoch. Individual ERPs were computed by averaging epochs in which the absolute amplitude at any sample was below 90 µV, at any electrode site. This procedure led to the rejection of approximately 39% of epochs (the rejected artefacts were mainly due to movement of the sleeping newborns and the associated slight shifts in the position of the electrodes, which can modulate the polarisation voltage, leading to changes in the recorded signals; some artefacts were probably also due to transitions between sleep stages, eye and involuntary muscle movements in active sleep). Table 1 displays the mean number and range of retained epochs, aggregated across infants and channels. Furthermore, the ERPs were subjected to offline digital filtering using a low-pass Savitzky-Golay filter (Press et al. 1992) with a first polynomial order and a window of 21 samples. This filtering enhanced the legibility of the responses.

2.5 Statistical analysis

Onset ERPs and offset ERPs were computed for standard stimuli in the spectral and duration conditions, respectively, excluding the two standards immediately following a deviant. Onset ERP was quantified as the area under curve in a window between 150 ms and 400 ms after vowel onset: in order to assess the ERP response related to the spectral difference that sets on at vowel onset, the window between 150 ms and 400 ms was intended to capture first ERP peak that in young (incl. premature) infants reportedly has a latency of about 200-250 ms and is considerably wider than adult ERPs (Eggermont & Moore 2011). Offset ERP was quantified as the area under curve in a window between 400 and 650 ms after vowel onset: in order to quantify the ERP response related to the durational difference between the short and the long vowel, offset ERP was assessed in a 250-ms window starting 220 ms after the offset of the short vowel.

Difference waves were calculated for physically identical stimuli, whereby the ERPs to standards from one block were subtracted from the ERPs to deviants – physically identical stimuli as the standards – from another block. MMR was calculated as area under curve in two time windows of the difference wave: an early window 80 to 220 ms after change onset and a late window 500 to 700 ms after change onset; the change onset coincided with vowel onset for the spectral change between [ɛ] and [a], and with the end of the short vowel for the duration change between [ɛ] and [ɛː].

Onset ERPs, offset ERPs, and the MMR were analyzed with linear mixed-effects models. The analysis for onset ERP modelled Conception age (continuous numeric factor, centered to 259 days, i.e. 37 weeks, considered as the threshold of term age), Stimulus (*a* vs *e*, coded as -1 vs +1), Region (lateral sites F3, C3, F4, C4 vs midline sites Fz, Cz, as -1 vs +1), and their interactions, and Sex (female vs male, coded as -1 vs +1) as fixed factors, and per-participant intercept and slopes for Stimulus and Region as random factors. The analysis for offset ERP modelled Conception age (numeric, centered to 259 days), Stimulus (*long ee* vs *short e*, coded as -1 vs +1), Region (lateral sites F3, C3, F4, C4 vs midline sites Fz, Cz, as -1 vs +1), and their interactions, and Sex (female vs male, coded as -1 vs +1) as fixed factors, and per-participant intercept and slopes for Stimulus and Region as random factors. The model for MMR included Conception age (continuous numeric factor, centered to 259 days), Contrast (durational vs. spectral, coded as -1 vs +1), Window of analysis (early vs late, coded as -1 vs +1), Direction of change (a change from [ɛ] to [ɛː] or [a] coded as -1, vs a change towards [ɛ] from [ɛː] or [a] coded as +1), and Region (lateral vs midline, coded -1 vs +1) as well as their interactions, and a main effect of Sex (F vs. M), as fixed factors, and a per-participant random intercept. The models were run in R (R Core team 2022) using the packages *lmer* and *lmerTest* (Bates et al. 2015, Kuznetsova et al. 2017), means were estimated with *ggeffects* (Lüdecke 2018).

3 Results

3.1 ERP results

Figure 3 plots the ERP waveform to Standard stimuli in each condition. The ERPs in the Spectral condition were statistically analyzed with the Onset models, and the ERPs in the Duration condition with the Offset models. The fixed-effects model summaries for Onset and Offset ERPs are shown in Table 2 and Table 3, respectively.

In the model for Onset ERP, the significant intercept indicates that overall there was an onset response reliably different from 0, with mean area under curve estimated at 201 µV·ms. There was also a main effect of Age, showing that the higher the Conception age the larger the peak. Inspection of the estimated means shows that the onset ERP amplitude was reliably larger than 0 from day 253 of conception age. There were also significant main effects of Stimulus and Region, suggesting that the Onset response was larger for the standard [ɛ] than for the standard [a], and larger on the midline than laterally. There were also significant two-way interactions of Age and Stimulus, and of Age and Region. The interaction of Age and Stimulus is directly relevant to our research question: "At what age does the ERPs reflect differential processing of different native vowels?". The interaction is visualised in Figure 4 (left), which indicates that the Onset ERPs to [a] and the Onset ERPs to [ɛ] start to differ from one another with increasing age. Inspections of the estimated means across the age range show that the onset ERPs to [a] and [ɛ] differ reliably from the 258th day of conception (pooled across the midline and lateral regions).

The model for Offset ERP did not yield a significant intercept, suggesting that overall no reliable Offset response was detected across conditions. There was a significant interaction of Age and Stimulus. As shown in Figure 4 (right), the Offset ERPs to [ɛː] and the Offset ERPs to [ɛ] start to differ from one another with increasing age. Inspections of the estimated means across the age range show that the Offset ERPs to [ɛː] and [ɛ] differ reliably from the 265th day of conception (pooled across the midline and lateral regions).

3.2 MMR results

The difference waveforms are shown in Figure 5. The fixed-effects model summary for MMR is shown in Table 4. As per the non-significant intercept, the analyses found no evidence of a reliable MMR across ages and conditions. However, Age was found to interact with the Window of analysis and with Contrast (Age:Latency: mean slope = -2.443, *p* = 0.010; Age:Dimension: mean slope = 4.197, *p* = 0.036). Figure 6 plots the MMR amplitude across the age range separately for each contrast and each window. It can be seen that the MMR amplitude gets more negative (supposedly indicating a more mature response) with increasing age, and especially so for the durational [ɛː]-[ɛ] contrast in the late MMR window. Inspection of the estimated means and their confidence intervals shows that a reliable MMR response is detected (only) for the late MMR to [ɛː]-[ɛ], which has a positive amplitude at the younger ages and becomes reliably negative (95% conf.int. below zero) at conception age 285 days.

4 Discussion

The present study sought to determine at what age in perinatal development, infants start to reliably discriminate between native vowels. Prior studies show that infants learn about the native language already *in utero*, being able to recognize previously exposed language sound patterns such as the language itself, its melody, rhythm, and very likely also individual speech segments or syllables (Moon et al. 2013, Partanen et al. 2013). At the same time, newborn infants process language stimuli differently from non-speech signals (May et al. 2018, Chládková et al. 2021). The early neural processing of speech thus very likely develops differently than the early neural processing of non-speech sounds. Yet, a fine- grained rajectory of early – prenatal or postnatal – development of auditory perception has almost exclusively been assessed with non-speech stimuli (Rotteveel et al 1987, Kushnerenko et al 2002, Bisiacchi et al. 2009, Lippé et al 2009, Suppiej et al. 2010). Studies that did use speech stimuli mostly compared preterm and fullterm infants at term age, thus not allowing to trace the very trajectory of the perinatal ERP development (Peña et al. 2012, François et al. 2019, Kostilainen et al 2020). The aim of the present experiment was to test at which age the brain starts to distinguish between acoustically different vowels (all of which belong to the phoneme inventory of the infants' native language). To this end, we recruited newborns in the age range between 32 and 42 weeks competition age (all tested a few days after birth) and measured their event-related potentials, as well as their auditory neural mismatch response, to native vowels differing in spectral quality and native vowels differing in duration.

A total of 99 sleeping newborns were played naturally produced, isolated vowels embedded in an oddball paradigm, half of the infants was tested with the vowels [ɛ] and [a] and the other half with the vowels [ɛ] and [ɛː]. Each infant listened to two oddball blocks, such that each of the two vowels served as a standard in one block and as a deviant in the other block. All these vowels represent phonemes in the infants' mothers' native language, Czech. We included both a spectral-change contrast (represented by the [ɛ] and [a]) as well as duration-change contrast (represented by the [ɛ] and [ɛː]) as we predicted that neural sensitivity might develop slightly earlier for the durational than for the spectral contrast. This prediction is based on the cue-specific properties of prenatal input, where durational cues are preserved in utero in an unchanged form, while spectral cues are modulated as they pass through tissues and amniotic water to the fetal ear (Richards et al. 1992, Granier-Deferre et al. 2011). Moreover, developmental studies with Czech-learning infants suggest more robust discrimination of vowel length compared to vowel quality across the first year of life, as well as exaggeration of durationally cued vowel contrasts in the infants' (prenatal) input (Chládková et al. 2019, Chládková et al. 2021. Paillereau et al. 2021, Svoboda et al. 2023). To assess the newborns' sensory processing of vowel acoustic properties we analysed the event-related potentials to vowel onsets and vowel offsets. To quantify the brain's neural discrimination of the vowel differences, we assessed the mismatch responses to the spectral change and to the durational change.

For the sensory responses to vowel onsets, the present analyses detected a positive peak from day 253 of conception age, that is, from 36 weeks and 1 day. This finding is in line with the literature demonstrating that in newborns the most prominent auditory ERP is a positive peak slightly after 200 ms, which with age develops into a negative N1 peak as the prominent auditory ERP response, maturing into the P1-N1-P2 complex (Pictor & Taylor 2007, Wunderlich et al. 2006). The present ERP results showed that the infants' onset ERP responses reflected the spectral differences between [ɛ] and [a] from conception age of 258 days (i.e., 36 weeks and 6 days), and their offset ERP responses reflected the durational difference between [ɛ] and [ɛː] from conception age 265 days (i.e. 37 weeks and 6 days). The onset response was more robust overall, which aligns with prior studies on adults that offset ERPs are smaller than onset ERPs (Baltzell & Billings 2014). A previous study comparing auditory onset and offset responses in young infants suggests that a large offset response is a marker of immature development (Wakai et al. 2007). The present study adds to that by showing that compared to onset responses, offset ERPs begin to reflect acoustic differences between vowels at a slightly later age (namely, one week later than onset ERPs). In the present study, onset responses reflected processing of the vowels' spectral characteristics and offset responses reflected processing of the vowels' durational characteristics. This might possibly indicate that the processing of spectral vowel contrasts matures earlier than the processing of vowel duration contrasts. However, in order to make conclusions about the order of maturation for the two types of contrasts, one would need to test them in a single model, ideally using a within-subject design, and – as far as possible – unconfounded by the positional context (i.e. whether the vowel change occurs at stimulus onset versus offset).

The MMR data showed that the newborns' brains discriminated the change in vowel duration, i.e. discriminated the speech contrast represented by [ɛ]–[ɛː], while no evidence of discrimination was found for vowel spectral quality. Interestingly, the MMR polarity inversely correlated with conception age, it was positive in the youngest infants and negative for the oldest infants. Although this developmental polarity change would align with some prior studies claiming a developmentally-conditioned MMR polarity (see Govaart, Dvořáková, et al. 2023 for a review), it is questionable to what extent one can validly assess an MMR response in the absence of reliably different sensory ERPs (Kremláček et al. 2016). We thus make no further inferences here regarding the MMR detected in infants younger than 253 days of conception age, since it was from this age when the vowel stimuli elicited a sensory ERP reliably different from zero. Considering the newborns older than 253 days, an MMR reliably different from 0, here with negative polarity, was elicited from conception age 285 days, that is, 40 weeks and 4 days. The presence of an MMR for the durational contrast (and the failure to detect it for the spectral contrast) aligns with prior studies documenting more a robust perceptual sensitivity to vowel duration of vowel spectrum in Czech-learning infants (Chládková et al. 2021, Paillereau et al. 2021).

The question remains whether the early maturation of an MMR response to vowel duration specifically (and its lack for vowel spectral changes) is language-specific, dependent on the infants' early prenatal input, or whether it is a language-general property of the developing speech perception system. To this end, a comparison to an earlier study with Finnish-learning newborns seems to speak in favour of language-specific MMR patterns at birth as the study with Finnish newborns found an MMR both for durational and spectral vowel contrasts, and detected an MMR in both an early and a late time window. This is, however, only a very rough comparison, since the two studies used very different types of stimuli (isolated vowels here and disyllabic words in the Finnish study) and different recording procedures and analysis pipelines. Note that the present study was not designed to test language-specific versus language-universal newborn speech perception and the present results cannot be interpreted in terms of language-specific phonological category learning in the perinatal period. The present results track the perinatal development of neural processing of different types of vowels (all of which happen to be realisations of phonemes in the ambient language).

The present findings contribute a more detailed understanding of the developmental trajectory of speech perception development in the perinatal period. Firstly, the present finding that reliable onset ERPs were elicited from conception age 36 weeks and 1 day aligns well with maturation of auditory brainstem potentials that were reported to stabilise at gestation week 36 (Starr et al. 1977). Secondly, as to the differentiation of acoustically distinct speech sounds, we identified 36 weeks and 6 days after conception as the age from which the cortex of (Czech-learning) newborns differentiates (at least some of) the native vowel identities. Since the ERPs are locked to the very occurrence of the target phonetic property, the present findings thus demonstrate a temporally rather precise phonetic perception three weeks prior to term age. Note that prior research has indicated that the newborn cortex is able to differentiate between strings of [ga] and [ba] syllables already before 35 weeks of gestation age (assessing blood oxygenation levels in Mahmoudzadeh et al. 2013, and ERPs in Mahmoudzadeh et al. 2017). Compared to Mahmoudzadeh et al. (2017), the later onset of reliable ERPs in our experiment might be due to different stimulus identities and presentation paradigms (blocks of syllables interspersed by silences in the previous study vs. trains of vowels in the present study), different intensity levels (70 dB in the previous study vs 65 dB in the present study), or different procedures and equipment. The present findings of precise phonetic perception just prior to term age extend our knowledge on the capacities of the newborn brain to learn the ambient speech sounds. Using EEG and fNIRS, previous studies documented fast phonetic learning for previously unexposed vowels in full-term newborns and two-month old infants (Cheour et al. 2002, Wanrooij et al. 2014, Wu et al. 2022). Our findings of reliably differentiated ERPs from conception age 36 weeks and 6 days demonstrate that accurate phonetic perception of contrastive vowel properties is in place before (or at least at the same time as) the age at which infants have been reported to learn novel vowels from exposure.

The age at which ERPs start to reliably distinguish between acoustically different vowels seems to rather well coincide with the age that is, in many countries considered as the term age. In that respect, ERPs to vowel onsets and offsets might have the potential to help identify children with developmental delays, particularly those pertaining to speech and language, such as dyslexia. Hämäläinen et al. (2013) showed that atypical ERPs to speech and nonspeech sounds in preschool children are related to poorer reading abilities at school age. Atypical speech processing seems to index dyslexia already at birth in that full term newborns with familial risk of dyslexia reportedly showed delayed, attenuated, or even lacking MMR responses to vowel changes in disyllabic words (Thiede et al. 2019). Since the occurrence of primary ERPs reliably distinguishing phonetically different vowels coincides here quite well with term age, i.e. the age of maturation, recording the primary ERPs to isolated vowel sounds might prove as a suitable method for assessing developmental language delays at birth. Future research is needed to collect normative data on ERPs to isolated speech sounds at birth (for the language community at hand) and test whether newborns whose ERPs deviate from the norm develop speech or language pathologies later in life. If that is the case, the early identification of potential language delays would allow targeting a focused therapy (more speech input, more systematic input. etc.) from the earliest possible age.

Acknowledgments

This experiment was carried out at the Neonatal unit of the Paediatric Department at Havlíčkův Brod Hospital, Czech Republic, where the first author is a practising paediatrician. We are very grateful to Havlíčkův Brod hospital and, in particular, the staff at the maternity ward and the neonatal department for accommodating our experiment. This work was supported by the European Regional Development Fund (ERDF) project "Brain dynamics" (reg. no.: CZ.02.01.01/00/22_008/0004643), by the ERDF project “Beyond Security: Role of Conflict in Resilience-Building” (reg. no.: CZ.02.01.01/00/22_008/0004595), and by the Czech Science Foundation (grant no. 21-09797S).

References

- Abboub, N., Nazzi, T., & Gervain, J. (2016). Prosodic grouping at birth. Brain and language, 162, 46-59.
- Baltzell, L. S., & Billings, C. J. (2014). Sensitivity of offset and onset cortical auditory evoked potentials to signals in noise. Clinical neurophysiology: official journal of the International Federation of Clinical Neurophysiology, 125(2), 370.
- Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). “Fitting Linear Mixed-Effects Models Using lme4.” Journal of Statistical Software, 67(1), 1–48.
- Bisiacchi, P. S., Mento, G., & Suppiej, A. (2009). Cortical auditory processing in preterm newborns: an ERP study. Biological psychology, 82(2), 176-185.
- Boersma, P., & Weenink, D. (2024). Praat: doing phonetics by computer [Computer program]. Version 6.4.05, retrieved 27 January 2024 from http://www.praat.org/
- Cheour, M., Martynova, O., Näätänen, R., Erkkola, R., Sillanpää, M., Kero, P., ... & Hämäläinen, H. (2002). Speech sounds learned by sleeping newborns. *Nature*, 415(6872), 599-600.
- Chládková, K., Černá, M., Paillereau, N., Skarnitzl, R., & Oceláková, Z. (2019). Prenatal infant-directed speech: vowels and voice quality. Proceedings of the 19th ICPhS, Melbourne, pp. 1525-1529.
- Chládková, K., Urbanec, J., Skálová, S., & Kremláček, J. (2021). Newborns’ neural processing of native vowels reveals directional asymmetries. Developmental Cognitive Neuroscience, 52, 101023.
- Daneshvarfard, F., Abrishami Moghaddam, H., Dehaene-Lambertz, G., Kongolo, G., Wallois, F., & Mahmoudzadeh, M. (2019). Neurodevelopment and asymmetry of auditory-related responses to repetitive syllabic stimuli in preterm neonates based on frequency-domain analysis. Scientific Reports, 9(1), 10654.
- DeCasper, A. J., & Fifer, W. P. (1980). Of human bonding: Newborns prefer their mothers' voices. Science, 208(4448), 1174-1176.

- DeCasper, A. J., Lecanuet, J. P., Busnel, M. C., Granier-Deferre, C., & Maugeais, R. (1994). Fetal reactions to recurrent maternal speech. Infant behavior and development, 17(2), 159-164.
- Delorme, A., & Makeig, S. (2004). EEGLAB: an open source toolbox for analysis of singletrial EEG dynamics including independent component analysis. Journal of neuroscience methods, 134(1), 9-21.
- Eggermont, J. J., & Moore, J. K. (2011). Morphological and functional development of the auditory nervous system. In Human auditory development (pp. 61-105). New York, NY: Springer New York.
- François, C., Rodriguez- Fornells, A., Teixidó, M., Agut, T., & Bosch, L. (2021). Attenuated brain responses to speech sounds in moderate preterm infants at term age. Developmental Science, 24(1), e12990.
- Garrido, M. I., Kilner, J. M., Stephan, K. E., & Friston, K. J. (2009). The mismatch negativity: a review of underlying mechanisms. Clinical neurophysiology, 120(3), 453-463.
- Govaart, G. H., Dvořáková, M., Chládková, K., & Männel, C. (2023). Infant Brain Responses in Auditory Perception: A Review of the Factors Influencing the Polarity of the Mismatch Response. Preprint at https://osf.io/bx952/.
- Granier- Deferre, C., Ribeiro, A., Jacquet, A. Y., & Bassereau, S. (2011). Near- term foetuses process temporal features of speech. Developmental science, 14(2), 336-352.
- Guzzetta, F., Conti, G., & Mercuri, E. (2011). Auditory processing in infancy: do early abnormalities predict disorders of language and cognitive development? Developmental Medicine & Child Neurology, 53(12), 1085-1090.
- Hämäläinen, J. A., Guttorm, T. K., Richardson, U., Alku, P., Lyytinen, H., & Leppänen, P. H. (2013). Auditory event-related potentials measured in kindergarten predict later reading problems at school age. Developmental neuropsychology, 38(8), 550-566.
- Huotilainen M, Kujala A, Hotakainen M, Parkkonen L, Taulu S, Simola J, et al. Short-term memory functions of the human fetus recorded with magnetoencephalography. Neuroreport. 2005 Jan 19;16(1):81–4.
- Kostilainen, K., Partanen, E., Mikkola, K., Wikström, V., Pakarinen, S., Fellman, V., & Huotilainen, M. (2020). Neural processing of changes in phonetic and emotional speech sounds and tones in preterm infants at term age. International Journal of Psychophysiology, 148, 111-118.
- Kremláček, J., Kreegipuu, K., Tales, A., Astikainen, P., Poldver, N., Näätänen, R., & Stefanics, G. (2016). Visual mismatch negativity (vMMN): A review and meta-analysis of studies in psychiatric and neurological disorders. Cortex, 80, 76-112.
- Kushnerenko, E., Ceponiene, R., Balan, P., Fellman, V., Huotilainen, M., & Näätänen, R. (2002). Maturation of the auditory event-related potentials during the first year of life. Neuroreport, 13(1), 47-51.
- Kuznetsova, A., Brockhoff, P.B., Christensen, R.H.B. (2017). “lmerTest Package: Tests in Linear Mixed Effects Models.” Journal of Statistical Software, 82(13), 1–26.
- Lecanuet, J. P., Granier-Deferre, C., DeCasper, A. J., Maugeais, R., Andrieu, A. J., & Busnel, M. C. (1987). Fetal perception and discrimination of speech stimuli; demonstration by cardiac reactivity; preliminary results. Comptes Rendus de l'academie des Sciences. Serie III, Sciences de la vie, 305(5), 161-164.
- Lippé, S., Kovacevic, N., & McIntosh, R. (2009). Differential maturation of brain signal complexity in the human auditory and visual system. Frontiers in human neuroscience, 3, 792.

- Lüdecke, D. (2018). “ggeffects: Tidy Data Frames of Marginal Effects from Regression Models.” Journal of Open Source Software, 3(26), 772.
- Mahajan, Y., & McArthur, G. (2012). Maturation of auditory event-related potentials across adolescence. Hearing research, 294(1-2), 82-94.
- Mahmoudzadeh, M., Dehaene-Lambertz, G., Fournier, M., Kongolo, G., Goudjil, S., Dubois, J., ... & Wallois, F. (2013). Syllabic discrimination in premature human infants prior to complete formation of cortical layers. Proceedings of the National Academy of Sciences, 110(12), 4846-4851.
- Mahmoudzadeh, M., Wallois, F., Kongolo, G., Goudjil, S., & Dehaene-Lambertz, G. (2017). Functional maps at the onset of auditory inputs in very early preterm human neonates. Cerebral Cortex, 27(4), 2500-2512.
- Maitre, N. L., Lambert, W. E., Aschner, J. L., & Key, A. P. (2013). Cortical speech sound differentiation in the neonatal intensive care unit predicts cognitive and language development in the first 2 years of life. Developmental Medicine & Child Neurology, 55(9), 834-839.
- Mampe, B., Friederici, A. D., Christophe, A., & Wermke, K. (2009). Newborns' cry melody is shaped by their native language. Current biology, 19(23), 1994-1997.
- May, L., Gervain, J., Carreiras, M., & Werker, J. F. (2018). The specificity of the neural response to speech at birth. Developmental science, 21(3), e12564.
- Moon, C., Lagercrantz, H., & Kuhl, P. K. (2013). Language experienced in utero affects vowel perception after birth: A two- country study. Acta paediatrica, 102(2), 156-160.
- Paillereau, N., Podlipský, V. J., Šimáčková, Š., Smolík, F., Oceláková, Z., & Chládková, K. (2021). Perceptual sensitivity to vowel quality and vowel length in the first year of life. JASA Express Letters, 1(2).
- Partanen, E., Kujala, T., Näätänen, R., Liitola, A., Sambeth, A., & Huotilainen, M. (2013). Learning-induced neural plasticity of speech processing before birth. Proceedings of the National Academy of Sciences, 110(37), 15145-15150.
- Pena, M., Werker, J. F., & Dehaene-Lambertz, G. (2012). Earlier speech exposure does not accelerate speech acquisition. Journal of neuroscience, 32(33), 11159-11163.
- Picton, T. W., & Taylor, M. J. (2007). Electrophysiological evaluation of human brain development. Developmental neuropsychology, 31(3), 249-278.
- Press W.H., Teukolsky S.A., Vetterling W.T., Flannery B.P. (1992). *Numerical Recipes in C: The Art of Scientific Computing.* Savitzky-Golay smoothing filters, pp. 650–655. Cambridge: Cambridge University Press.
- R Core Team (2022). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. https://www.R-project.org/.
- Richards, D. S., Frentzen, B., Gerhardt, K. J., McCann, M. E., & Abrams, R. M. (1992). Sound levels in the human uterus. Obstetrics and Gynecology, 80(2), 186-190.
- Rotteveel, J. J., De Graaf, R., Stegeman, D. F., Colon, E. J., & Visco, Y. M. (1987). The maturation of the central auditory conduction in preterm infants until three months post term. V. The auditory cortical response (ACR). Hearing research, 27(1), 95-110.
- Ruhnau, P., Herrmann, B., Maess, B., & Schröger, E. (2011). Maturation of obligatory auditory responses and their neural sources: evidence from EEG and MEG. Neuroimage, 58(2), 630-639.
- Starr, A., Amlie, R. N., Martin, W. H., & Sanders, S. (1977). Development of auditory function in newborn infants revealed by auditory brainstem potentials. Pediatrics, 60(6), 831-839.

- Stefanics, G., Háden, G., Huotilainen, M., Balázs, L., Sziller, I., Beke, A., ... & Winkler, I. (2007). Auditory temporal grouping in newborn infants. Psychophysiology, 44(5), 697-702.
- Suppiej, A., Mento, G., Zanardo, V., Franzoi, M., Battistella, P. A., Ermani, M., & Bisiacchi, P. S. (2010). Auditory processing during sleep in preterm infants: An event related potential study. Early human development, 86(12), 807-812.
- Svoboda, M., Chládková, K., Kocjančič Antolík, T., Paillereau, N., & Slížková, P. (2023). Vowel length in infant-directed speech: the realisation of short-long contrasts in Czech IDS. Proceedings of the 20th ICPhS, Prague, pp. 2363–2367.
- Thiede, A., Virtala, P., Ala-Kurikka, I., Partanen, E., Huotilainen, M., Mikkola, K., ... & Kujala, T. (2019). An extensive pattern of atypical neural speech-sound discrimination in newborns at risk of dyslexia. Clinical Neurophysiology, 130(5), 634-646.
- Tomé, D., Barbosa, F., Nowak, K., & Marques-Teixeira, J. (2015). The development of the N1 and N2 components in auditory oddball paradigms: a systematic review with narrative analysis and suggested normative values. Journal of neural transmission, 122, 375-391.
- Wakai, R. T., Lutter, W. J., Chen, M., & Maier, M. M. (2007). On and off magnetic auditory evoked responses in early infancy: A possible marker of brain immaturity. Clinical neurophysiology, 118(7), 1480-1487.
- Wanrooij, K., Boersma, P., & Van Zuijen, T. L. (2014). Fast phonetic learning occurs already in 2-to-3-month old infants: an ERP study. *Frontiers in Psychology*, 5, 77.
- Wu, Y. J., Hou, X., Peng, C., Yu, W., Oppenheim, G. M., Thierry, G., & Zhang, D. (2022). Rapid learning of a phonemic discrimination in the first hours of life. *Nature Human Behaviour*, 6(8), 1169-1179.
- Wunderlich, J. L., & Cone-Wesson, B. K. (2006). Maturation of CAEP in infants and children: a review. Hearing research, 212(1-2), 212-223.

Figure 1. The 99 participants in the present experiment. Points show individual participants; gestation age at birth = x axis, conception age at time of experiment = y axis; the distribution of females and males in colour; and assignment to stimulation groups indicated by different shapes. Dimension was a between-subject factor, dividing our total sample into two groups of n = 51, and n = 48, respectively for the stimulation with the spectral contrast and for the stimulation with the durational contrast. Conception age in days was modelled as a continuous factor in the present study.

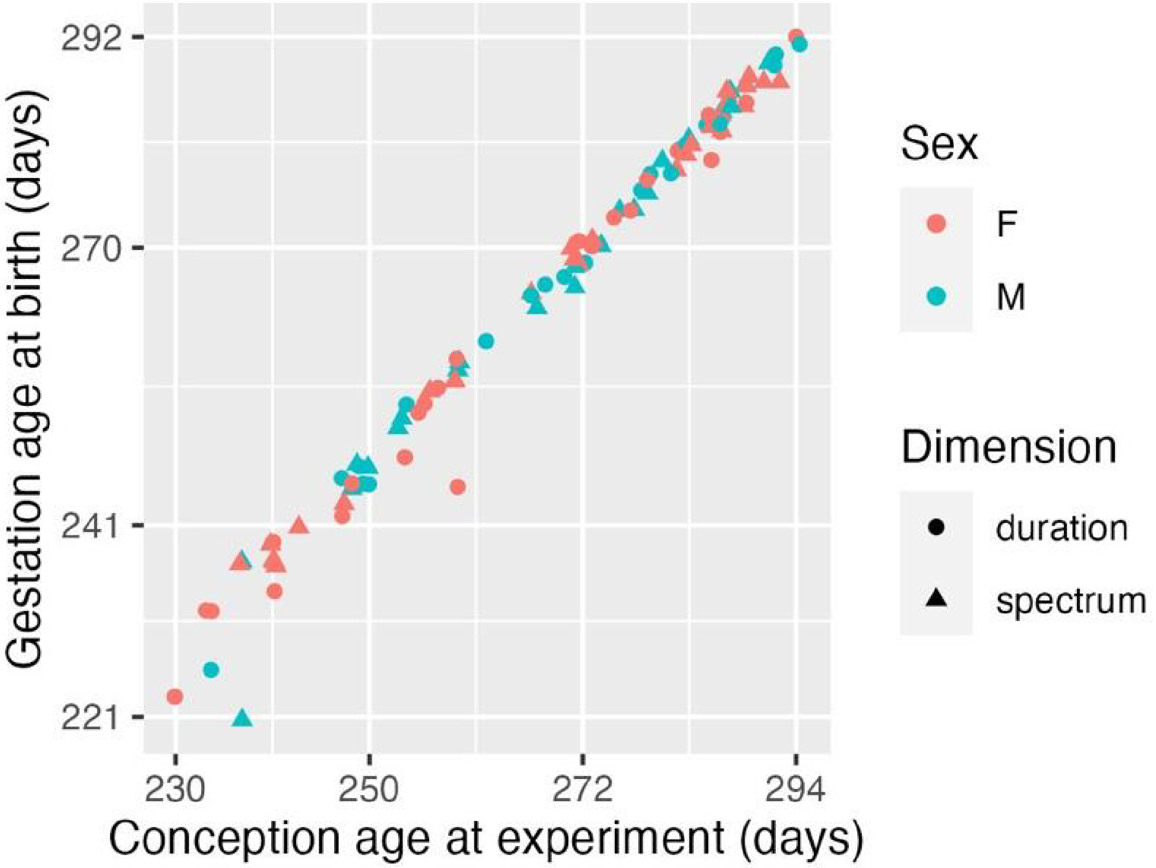


Figure 2. An asleep infant taking part in the experiment. The photo displays 2 of the 6 EEG sensors attached on the scalp (here, at locations F3 and C3), an external channel on the face (grounding), and one of the infant ear couplers with an insert earphone (the online-reference channel on the right side of the nose is not visible in this figure).

Figure 3. ERPs to standards in infants tested with the spectral change between [a] and [ɛ] (left) and in infants tested with the durational change between [ɛː] and [ɛ] (right). Individual rows show responses averaged across infants in one of four gestation age bins (the age bins are used only for visualisation; analyses were done with age as continuous factor). The figure shows averages for the lateral sites 12 (F3, C3, F4, C4) and the midline sites (Fz, Cz). Shaded areas represent 95% confidence intervals of the mean ERP waveforms.

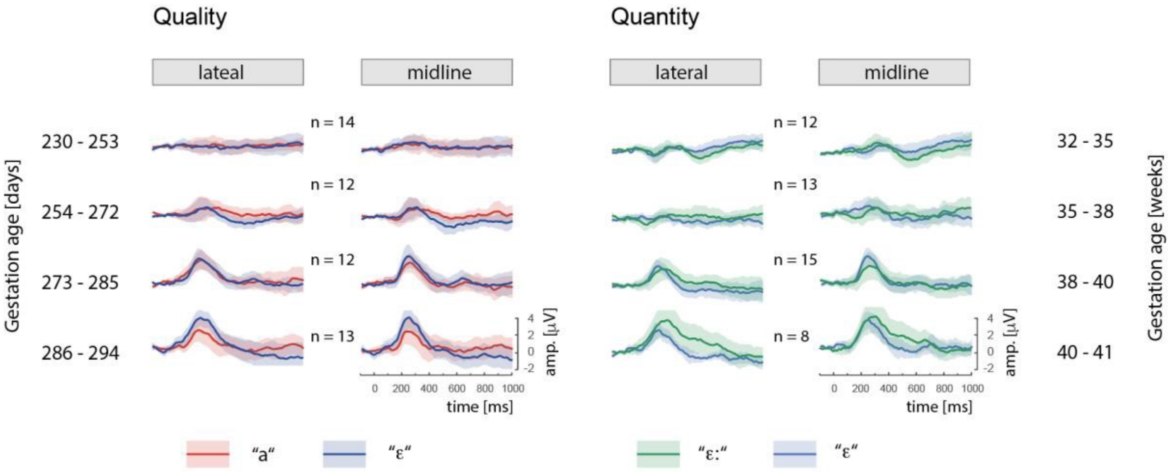


Figure 4. Modelled amplitude of the Onset ERP to standards in the spectral condition (left) and to Offset ERP to standards in the duration condition (right). The thick curves show estimated means and shaded sleeves represent 95% confidence intervals; points show raw data.

Figure 5. Left: Difference waves for the spectral contrast (*ε-a or a-ε*). Right: Difference waves for the durational contrast (*ε-εː or εː-ε)*. Individual rows show grand-average difference waves in four gestation age categories averaged over all recorded leads (F3, C3, F4, C4, CZ, FZ); the figure pools across both directions of change within each contrast type (for each direction of change differences waveforms were computed from physically-identical stimuli from different blocks and then averaged across the two directions of change). The black bars depict the time intervals used for the early (180–220 ms) and the late (500–700 ms) analysis window to compute the area under curve (AUC). Shaded areas represent 95% confidence intervals of the mean difference waveforms.

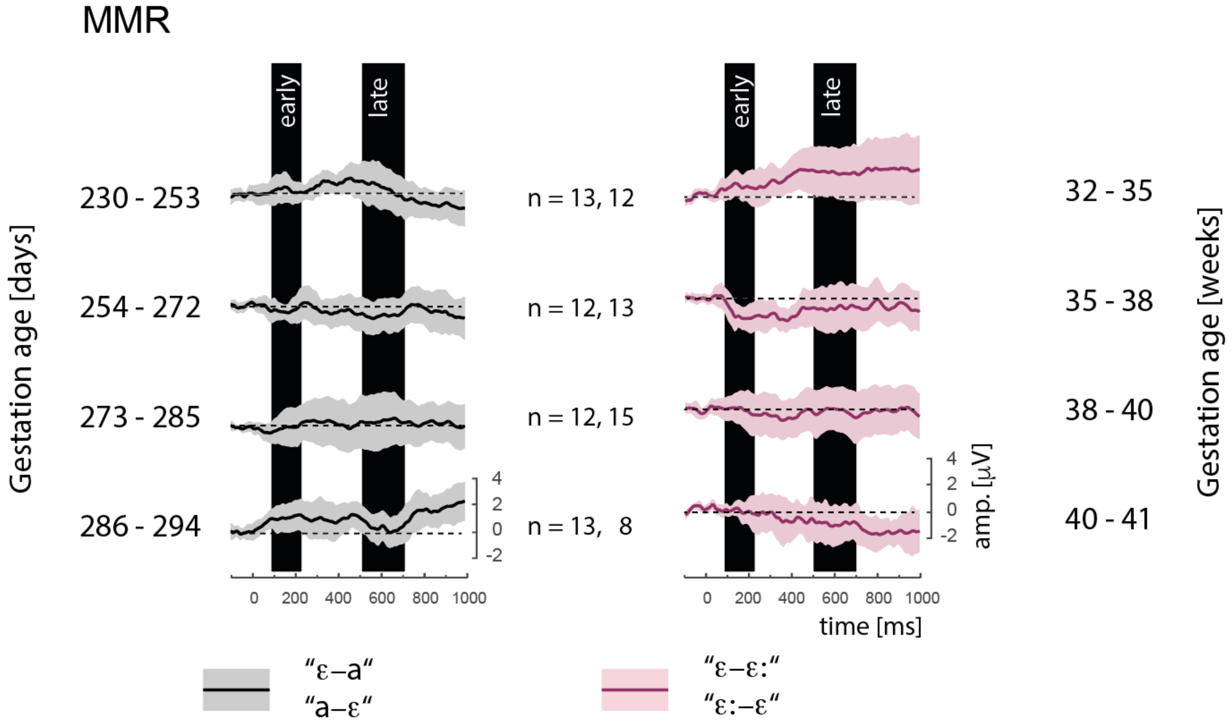


Figure 6. Modelled MMR per time window and per contrast, i.e. the durational (red) and the spectral change (black). Thick curves show estimated mean and shaded sleeves 95% confidence intervals; points show raw data.

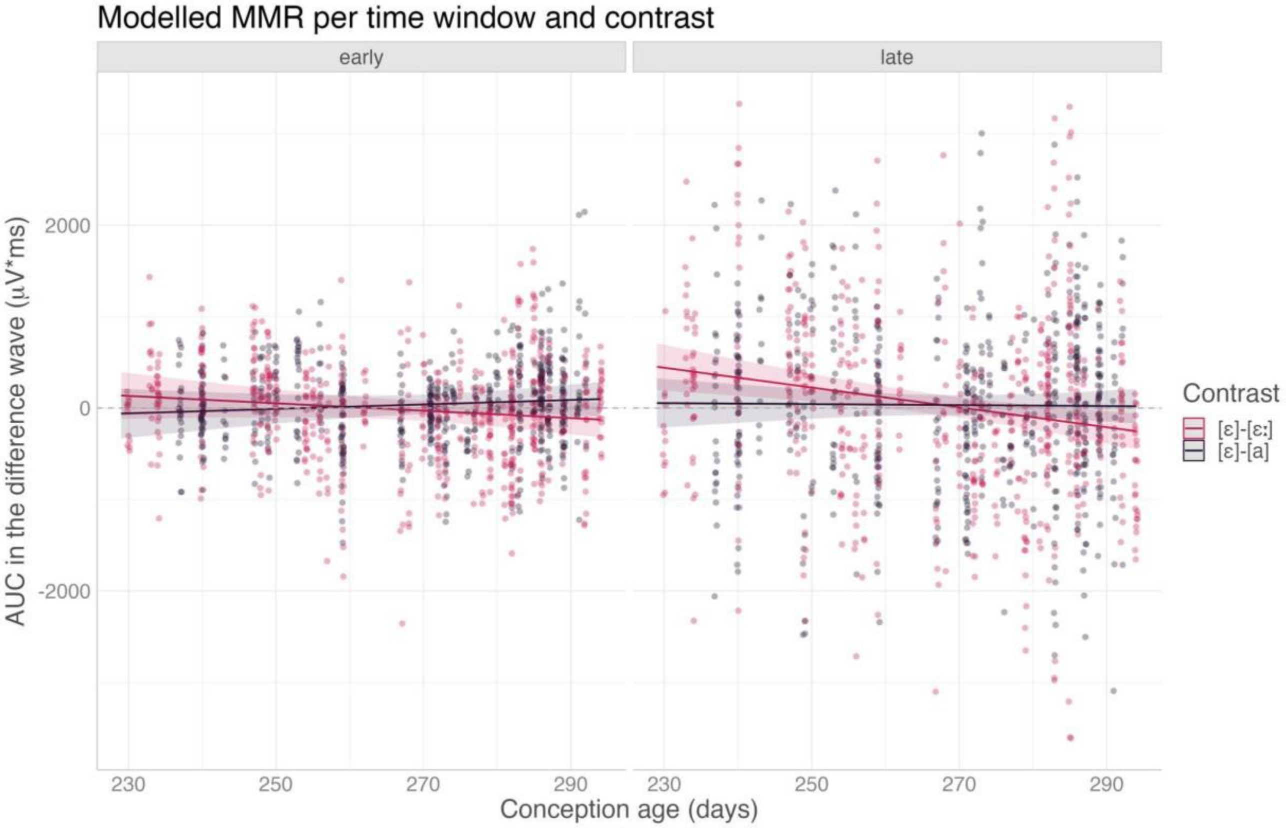


Table 1. The average, the minimum, and the maximum number of epochs pooled across infants (divided in two age bins for the purpose of the artefact rejection statistics in this Table only) and channels, for each stimulus type. The row for "ε(ε:)" represents the number of epochs for [ε] that were presented in the same session with [ε:] Analogously, the row for "ε(a)" represents the number of distinct epochs of [ε] played in the same sessions with [a].

		ERP standard			ERP deviant		
Age bin	Stimulus	min	mean	max	min	mean	max
Fullterm	ε:	89	291	438	17	71	119
	ε (ε:)	96	288	475	23	71	110
	ε (a)	61	311	473	18	75	114
	a	86	297	474	11	76	116
Preterm	ε:	90	288	461	20	77	112
	ε (ε:)	73	311	443	30	71	116
	ε (a)	126	271	440	26	68	111
	a	86	293	435	27	62	109

Table 2. Fixed-effects model output for onset ERP.

Parameters for ONSET ERP	Estimate	Std. Error	df	t value	Pr(>|t|)
(Intercept)	201.315	52.07	49.413	3.866	<0.001
Age (mean-centred)	12.097	2.592	49.36	4.668	<0.001
Stimulus (-a+ε)	-40.6	18.788	560.98	-2.161	0.031
Region (-lateral+midline)	37.32	18.784	560.935	1.987	0.047
Sex (-F+M)	69.728	45.673	47.925	1.527	0.133
Age:Stimulus	-1.917	0.934	561.191	-2.051	0.041
Age:Region	-2.122	0.933	560.935	-2.274	0.023
Stimulus:Region	-18.449	18.784	560.935	-0.982	0.326
Age:Stimulus:Region	0.289	0.933	560.935	0.309	0.757

Table 3. Fixed-effects model output for offset ERP.

Parameters for OFFSET ERP	Estimate	Std. Error	df	t value	Pr(>|t|)
(Intercept)	-19.849	42.427	47.656	-0.468	0.642
Age (mean-centred)	2.111	2.116	47.588	0.998	0.324
Stimulus (-e:+e)	15.628	21.412	522.000	0.730	0.466
Region (-lateral+midline)	-5.680	21.412	522.000	-0.265	0.791
Sex (-F+M)	-46.455	38.621	45.000	-1.203	0.235
Age:Stimulus	3.859	1.055	522.000	3.659	<0.001
Age:Region	0.340	1.055	522.000	0.322	0.747
Stimulus:Region	-4.868	21.412	522.000	-0.227	0.820
Age:Stimulus:Region	-1.225	1.055	522.000	-1.161	0.246

Table 4. Fixed-effects model output for MMR.

Parameters for MMR	Estimate	Std. Error	df	t value	Pr(>|t|)
(Intercept)	47.880	40.397	108.284	1.185	0.239
Age (mean-centred)	-3.241	1.991	104.439	-1.628	0.106
Dimension (-duration+spectrum)	-22.736	39.834	104.867	-0.571	0.569
MMR latency (-early+late)	34.155	19.257	2252.886	1.774	0.076
Direction of change (-from /E/, +to /E/)	17.255	19.258	2252.933	0.896	0.370
Region (-lateral+midline)	15.291	15.674	2252.886	0.976	0.329
Sex (-F+M)	37.877	34.746	92.868	1.090	0.276
Age:Dimension	4.197	1.979	104.504	2.120	0.036
Age:Latency	-2.443	0.952	2252.886	-2.565	0.010
Dimension:Latency	-21.286	19.257	2252.886	-1.105	0.269
Age:Direction	0.461	0.952	2253.205	0.484	0.628
Dimension:Direction	-23.118	19.258	2252.933	-1.200	0.230
Latency:Direction	-2.621	19.257	2252.886	-0.136	0.892
Age:Dimension:Latency	0.950	0.952	2252.886	0.998	0.319
Age:Dimension:Direction	0.866	0.952	2253.205	0.909	0.364
Age:Latency:Direction	0.512	0.952	2252.886	0.538	0.591
Dimension:Latency:Direction	-1.098	19.257	2252.886	-0.057	0.955
Age:Dimension:Latency:Direction	0.455	0.952	2252.886	0.478	0.633

6. ZÁVĚRY ORIGINÁLNÍCH PUBLIKACÍ

Sluchová percepce je komplexní a nenahraditelný proces získávání informací[5-8,10]. Zpracování zvukového signálu vyžaduje zapojení podkorových i vyšších, korových center. Publikované studie prokázaly prenatální stimulaci plodu zvukovými stimuly jako nezbytnou pro správný vývoj a postnatální vnímání zvuku[1,2]. Prenatální stimulací lze tuto percepci ovlivnit[17]. Porucha vývoje sluchové dráhy spojené s nedostatečnou zvukovou stimulací může vést u vyvíjejícího se plodu nebo dítěte k narušení správného psychomotorického vývoje. Objektivní vyšetření sluchové percepce by bylo možné využít k vyšetření CNS a případné diagnostice patologických stavů. K tomuto vyšetření jsou využívány kortikální sluchové potenciály. Tyto potenciály odráží vyspělou mozkovou aktivitu komplexně zpracovávající zvukové vjemy[12]. Při narušeném vývoji dítěte vykazují evokované potenciály abnormální hodnoty například sníženou amplitudu nebo delší latenci. V klinické, zahraniční praxi má své místo využití těchto potenciálů ke sledování dynamiky sluchové percepce po zavedení kochleárního implantátu[7]. Do budoucna se nabízí více možností využití kortikálních potenciálů ke sledování a případné diagnostice především funkčních poruch CNS[10]. Nevýhodou a jistým limitem této metody je individuální variabilita korových potenciálů. Zatímco pro vybavení krátko – a střednědobých potenciálů z podkorových struktur je možné využít univerzální, tónovou stimulaci, pro vyvolání evokovaných potenciálů při hodnocení sluchové percepce je nutná unikátní stimulace. Každý národní jazyk je charakterizován typickými hláskovými kontrasty. Tyto kontrasty jsou nutné pro dostatečnou zvukovou stimulaci a adekvátní sluchovou percepci. Výzkum a vývoj vhodné zvukové stimulace se jeví jako naprosto klíčový pro správné pochopení a následné využití metody sluchových evokovaných potenciálů v širší praxi.

Výsledky našeho výzkumu potvrdily, že u donošených novorozenců již krátce po narození je sluchová percepce vyzrálá a podobná dospělým jedincům. Zvolená zvuková stimulace vedla k vybavení dostatečné korové odpovědi. V souladu s dostupnými poznatky by tak bylo možné využít navrženou stimulaci k vyšetření novorozenců a ranému odhalení patologických stavů, který by mohly narušit harmonický psychomotorický vývoj. Prostorem pro další výzkum je získání ideální křivky kortikálních potenciálů, která by sloužila jako referenční křivka k posouzení případných poruch CNS.

V práci „Neural Processing of Spectral and Durational Changes in Speech and Non-speech Stimuli: An MMN Study With Czech Adults“ jsme vlastní, originálně připravenou stimulací s hláskovými a tónovými kontrasty vyvolali dostatečnou korovou odpověď u dospělých dobrovolníků. Tato odpověď byla následně vyhodnocena pomocí MMN. Největší amplitudy byly zachyceny v levostranné, frontální lokalizaci distribuce sluchové a lingvistické odpovědi

MMN. Získaná data potvrdila náš předpoklad o komplexním a vyzrálém zpracování zvuku nejen v primární sluchové kůře, ale především v asociovaných korových oblastech. Sluchová percepce a zpracování zvuku jsou tedy nezbytné pro kognitivní funkce CNS.

V práci publikované v článku „Newborns´ neural processing of native vowels reveals directional asymmetries“ jsme prokázali schopnost donošených novorozenců vnímat hláskové kontrasty českého jazyka. Překvapujícím zjištěním bylo vyzrálé a komplexní zpracování zvukové stimulace srovnatelné s dospělými jedinci. Tato odpověď byla také hodnocena pomocí MMR. MMR byla s negativní amplitudou a levostranně detekovatelná pro změnu délky trvání hlásek [fe] -> [fe:] i změnu spektrální kvality hlásek [fa] -> [fe]. Můžeme ji tak nazývat MMN. Negativní hodnoty pro amplitudu také svědčí pro vyspělou odpověď srovnatelnou s dospělými. Tónová stimulace tuto odpověď vyvolat nedokázala. Tento závěr potvrzuje nezbytnost adekvátní sluchové stimulace (především řečové) ke správnému vývoji dítěte a vyzrávání kognitivních funkcí[1,2].

V závěrečné práci „Neural processing of speech sounds at premature and term birth: ERPs and MMR between 32 and 42 weeks of gestation“ jsme prokázali postupné vyzrávání adekvátní sluchové percepce, které je přímo úměrné rostoucímu gestačnímu stáří. Od zahájeného 36. týdne gestace je vyvíjející se plod schopný rozpoznat spektrální kvalitu hlásek, od 37. týdne i jejich délkovou kvalitu. Zpracování hlásek lze hodnotit pomocí MMR, která může mít ještě pozitivní amplitudu. V termínu porodu byla opětovně zaznamená odpověď komplexně vyzrálá s negativní amplitudou MMR. Dobře se vyvíjející novorozenec je schopný zpracovat přicházející řečovou stimulaci komplexně jako dospělý jedinec. Nedokáže ale přiřadit význam zvuku, který jej stimuluje. Abnormální hodnoty MMR mohou poukazovat na funkční nebo strukturální narušení vyvíjejícího se CNS.

7. SEZNAM OBRÁZKŮ

7.1 Součást práce „The development and neurophysiological assessment of newborn auditory cognition: a review of findings and their application“

7.2 Součást originální práce („Neural Processing of Spectral and Durational Changes in Speech and Non-speech Stimuli: An MMN Study With Czech Adults“

7.3 Součást originální práce „Newborns´ neural processing of native vowels reveals directional asymmetries“

Figure 4. Modelled means and 95% CIs for the onset ERP (left) depicting the interaction of Domain, Spectrum, and Anteriority, and for the offset ERP (right) depicting the interaction of Domain, Duration, and Laterality. Colour coding aligns with the colours of the grand average ERP waves plotted in Fig. 3.

Figure 5. Grand average difference waves in the three scalp regions (for region visualization, see Fig. 2). Shading shows the early and late MMR analysis windows. Numbers in the top right corners show over how many participants averaging was done in each condition. The difference waves were computed from physically identical stimuli, e.g. the difference wave for the spectral "from-E" deviant was computed as: ERP for [a] as deviant minus ERP for [a] as standard, and the difference wave for the spectral "to-E" deviant was computed as: ERP for [ɛ] as deviant minus ERP for [ɛ] as standard, and likewise for the durational deviations between [ɛ] and [ɛː].

Figure 6. The modelled means and 95% c.i.s for the early MMR, unpacking the Domain * Dimension * Deviant * Laterality interaction. Deviant label "to-E" corresponds to the [ɛ] deviant on the spectral dimension and to the short deviant on the duration dimension, and deviant label "from-E" corresponds to the [a] deviant on the spectral dimension and to the long deviant on the duration dimension.

7.4 Součást originální práce „Neural processing of speech sounds at premature and term birth: ERPs and MMR between 32 and 42 weeks of gestation"

Figure 1. The 99 participants in the present experiment. Points show individual participants; gestation age at birth = x axis, conception age at time of experiment = y axis; the distribution of females and males in colour; and assignment to stimulation groups indicated by different shapes. Dimension was a between-subject factor, dividing our total sample into two groups of n = 51, and n = 48, respectively for the stimulation with the spectral contrast and for the stimulation with the durational contrast. Conception age in days was modelled as a continuous factor in the present study.

Figure 2. An asleep infant taking part in the experiment. The photo displays 2 of the 6 EEG sensors attached on the scalp (here, at locations F3 and C3), an external channel on the face (grounding), and one of the infant ear couplers with an insert earphone (the online-reference channel on the right side of the nose is not visible in this figure).

Figure 3. ERPs to standards in infants tested with the spectral change between [a] and [ɛ] (left) and in infants tested with the durational change between [ɛː] and [ɛ] (right). Individual rows show responses averaged across infants in one of four gestation age bins (the age bins are used only for visualisation; analyses were done with age as continuous factor). The figure shows averages for the lateral sites 12 (F3, C3, F4, C4) and the midline sites (Fz, Cz). Shaded areas represent 95% confidence intervals of the mean ERP waveforms.

Figure 4. Modelled amplitude of the Onset ERP to standards in the spectral condition (left) and to Offset ERP to standards in the duration condition (right). The thick curves show

estimated means and shaded sleeves represent 95% confidence intervals; points show raw data.

Figure 5. Left: Difference waves for the spectral contrast (*ɛ-a or a-ɛ*). Right: Difference waves for the durational contrast (*ɛ-ɛː or ɛː-ɛ)*. Individual rows show grand-average difference waves in four gestation age categories averaged over all recorded leads (F3, C3, F4, C4, CZ, FZ); the figure pools across both directions of change within each contrast type (for each direction of change differences waveforms were computed from physically-identical stimuli from different blocks and then averaged across the two directions of change). The black bars depict the time intervals used for the early (180–220 ms) and the late (500–700 ms) analysis window to compute the area under curve (AUC). Shaded areas represent 95% confidence intervals of the mean difference waveforms.

Figure 6. Modelled MMR per time window and per contrast, i.e. the durational (red) and the spectral change (black). Thick curves show estimated mean and shaded sleeves 95% confidence intervals; points show raw data.

8. SEZNAM TABULEK

8.1 Součást originální práce („Neural Processing of Spectral and Durational Changes in Speech and Non-speech Stimuli: An MMN Study With Czech Adults“

8.2 Součást originální práce „Newborns´ neural processing of native vowels reveals directional asymmetries“

Table 6. Modelled means and standard errors (SE) for early and late MMR at left, midline, and right channels. Significance of pairwise comparisons (p.c.) across Deviants is indicated by asterisks: ** marks mutually exclusive means in the 95% confidence intervals estimated for each deviant type, * marks mutually exclusive means in 90% confidence intervals. Calculation of confidence intervals: 95% c.i. = mean ± 1.96SE, 90% c.i. = mean ± 1.645SE. The means and SEs were estimated using the ggeffects R package (Lüdecke, 2018, function ggpredict).

8.3 Součást originální práce „Neural processing of speech sounds at premature and term birth: ERPs and MMR between 32 and 42 weeks of gestation“

Table 1. The average, the minimum, and the maximum number of epochs pooled across infants (divided in two age bins for the purpose of the artefact rejection statistics in this Table only) and channels, for each stimulus type. The row for "ɛ(ɛː)" represents the number of epochs for [ɛ] that were presented in the same session with [ɛː] Analogously, the row for "ɛ(a)" represents the number of distinct epochs of [ɛ] played in the same sessions with [a].

Table 2. Fixed-effects model output for onset ERP.

Table 3. Fixed-effects model output for offset ERP.

Table 4. Fixed-effects model output for MMR.

9. LITERATURA

1. Graven S, Browne J. Auditory Development in the Fetus and Infant. NbInfant Nurs Rev 2008; **8**(4), 187–193.
2. Sousa A, Didoné D, Sleifer P. Longitudinal Comparison of Auditory Steady-State Evoked Potentials in Preterm and Term Infants: The Maturation Process. Int Arch Otorhinolaryngol 2017; **21**(03), 200-205.
3. Granier-Deferre C, Ribeiro A, Jacquet AY, et al. Near-term fetuses process temporal features of speech. Dev Sci 2011; **14**(2), 336-352.
4. Wodicka G, M. Lam A, Bhargava V, et al. Acoustic impedance of the maternal abdomen. J Acoust Soc Am 1993; **94**(1), 13-18.
5. Melo de A, Biaggio E, Rechia I, et al. Cortical auditory evoked potentials in full-term and preterm neonates. Codas 2016; **28**(5), 491-496.
6. Martins K, Gil D. Cortical Auditory Evoked Potentials with Simple (Tone Burst) and Complex (Speech) Stimuli in Children with Cochlear Implant. Int Arc Otorhinolaryngol 2017; **21**(04), 351-357.
7. Silva L, Couto M, Tsuji R, et al. Auditory pathways' maturation after cochlear implant via cortical auditory evoked potentials. Braz J Otorhinolaryngol 2014; **80**(2), 131-137.
8. Mehta K, Watkin P, Baldwin M, et al. Role of Cortical Auditory Evoked Potentials in Reducing the Age at Hearing Aid Fitting in Children With Hearing Loss Identified by Newborn Hearing Screening. *Trends in Hearing* 2017; **21**.
9. Oliveira L, Didoné D, Durante A. Automated cortical auditory evoked potentials threshold estimation in neonates. Braz J Otorhinolaryngol 2019; **85**(2), 206-212.
10. Thiede A, Virtala P, Ala-Kurikka I, et al. An extensive pattern of atypical neural speech-sound discrimination in newborns at risk of dyslexia. Clin Neurophysiol 2019; **130**(5), 634-646.
11. Frizzo, A. Auditory evoked potential: a proposal for further evaluation in children with learning disabilities. Front Psychol 2015; **6**, 788.
12. Duncan C, Barry R, Connolly J, et al. Event-related potentials in clinical research: Guidelines for eliciting, recording, and quantifying mismatch negativity, P300, and N400. Clin Neurophysiol 2009; **120**(11), 1883-1908.
13. Fellman V, Huotilainen M. Cortical auditory event-related potentials in newborn infants. Semin Fetal Neonatal Med 2006; **11**(6), 452-458.
14. Joos K, Gilles A, Van de Heyning P, et al. From sensation to percept: The neural signature of auditory event-related potentials. Neurosci Biobehav R 2014; **42**, 148-156.
15. Näätänen R, Picton T. The N1 wave of the human electric and magnetic response to sound: a review and an analysis of the component structure. Psychophysiology 1987; **24**(4), 375-425.
16. Lahav A, Skoe E. An acoustic gap between the NICU and womb: a potential risk for compromised neuroplasticity of the auditory system in preterm infants. Front Neurosci 2014; **8**, 381.
17. Partanen E, Kujala T, Näätänen R, Liitola A, Sambeth A, Huotilainen M. Learning-induced neural plasticity of speech processing before birth. Proc Natl Acad Sci U S A. 2013 Sep 10;110(37):15145-50.

Seznam použitých zdrojů je dále uveden pod jednotlivými originálními pracemi.

10. SOUHRN POZNATKŮ DISERTAČNÍ PRÁCE

Naše výsledky potvrdily vyzrálou sluchovou percepci donošených novorozenců. I když nejsou schopni posoudit význam jednotlivých slov, umí je rozlišit podle jejich hláskových kontrastů. Sluchová percepce je tak nezbytná pro správné fungování komplexních kognitivních funkcí. Toto zjištění je v souladu s předchozími studiemi. Výsledky našeho výzkumu jsou důležité pro detailní pochopení vývoje dětského mozku a jeho kognitivních funkcí. Rádi bychom je dále rozvíjeli a aplikovali v klinické praxi, kde se nabízí jejich široké využití např. v rámci diagnostiky funkčních poruch CNS.

Jméno a podpis studenta:	Jméno a podpis školitele:
MUDr. Josef Urbanec	Doc. MUDr. Sylva Skálová, Ph.D.