



# Early Development of Neural Speech Encoding Depends on Age but Not Native Language Status: Evidence From Lexical Tone

Nikolay Novitskiy<sup>1</sup> , Akshay R. Maggu<sup>1,2</sup> , Ching Man Lai<sup>1</sup> ,  
Peggy H. Y. Chan<sup>1,3</sup> , Kay H. Y. Wong<sup>1</sup> , Hugh Simon Lam<sup>3</sup> ,  
Tak Yeung Leung<sup>4</sup>, Ting Fan Leung<sup>3</sup>, and Patrick C. M. Wong<sup>1</sup>

<sup>1</sup>Department of Linguistics and Modern Languages, Brain and Mind Institute, The Chinese University of Hong Kong, Hong Kong SAR, China

<sup>2</sup>O-lab, Duke Psychology and Neuroscience, Duke University, Durham, NC, USA

<sup>3</sup>Department of Paediatrics, The Chinese University of Hong Kong, Hong Kong SAR, China

<sup>4</sup>Department of Obstetrics and Gynaecology, The Chinese University of Hong Kong, Hong Kong SAR, China

**Keywords:** infants, tone language, electroencephalography, frequency-following response, perceptual narrowing

## ABSTRACT

We investigated the development of early-latency and long-latency brain responses to native and non-native speech to shed light on the neurophysiological underpinnings of perceptual narrowing and early language development. Specifically, we postulated a two-level process to explain the decrease in sensitivity to non-native phonemes toward the end of infancy. Neurons at the earlier stages of the ascending auditory pathway mature rapidly during infancy facilitating the encoding of both native and non-native sounds. This growth enables neurons at the later stages of the auditory pathway to assign phonological status to speech according to the infant's native language environment. To test this hypothesis, we collected early-latency and long-latency neural responses to native and non-native lexical tones from 85 Cantonese-learning children aged between 23 days and 24 months, 16 days. As expected, a broad range of presumably subcortical early-latency neural encoding measures grew rapidly and substantially during the first two years for both native and non-native tones. By contrast, long-latency cortical electrophysiological changes occurred on a much slower scale and showed sensitivity to nativeness at around six months. Our study provided a comprehensive understanding of early language development by revealing the complementary roles of earlier and later stages of speech processing in the developing brain.

## INTRODUCTION

During early infancy, infants with normal hearing are able to perceive a large inventory of speech sounds that are both within and outside of their native language environments. Toward the end of infancy, this broad perceptual ability is gradually reduced to a smaller inventory of sounds that are mostly confined to the infants' native language environments (Werker & Tees, 1984). The term *perceptual narrowing* or *attunement* has been used to describe this phenomenon, and it is generally viewed as the basis of early language acquisition

Citation: Novitskiy, N., Maggu, A. R., Lai, C. M., Chan, P. H. Y., Wong, K. H. Y., Lam, H. S., Leung, T. Y., Leung, T. F., & Wong, P. C. M. (2022). Early development of neural speech encoding depends on age but not native language status: Evidence from lexical tone. *Neurobiology of Language*, 3(1), 67–86. [https://doi.org/10.1162/nol\\_a\\_00049](https://doi.org/10.1162/nol_a_00049)

DOI:  
[https://doi.org/10.1162/nol\\_a\\_00049](https://doi.org/10.1162/nol_a_00049)

Supporting Information:  
[https://doi.org/10.1162/nol\\_a\\_00049](https://doi.org/10.1162/nol_a_00049)

Received: 21 October 2020  
Accepted: 22 June 2021

Competing Interests: The authors declare competing interests.

Corresponding Author:  
Patrick Wong  
[p.wong@cuhk.edu.hk](mailto:p.wong@cuhk.edu.hk)

Handling Editor:  
Janet F. Werker

Copyright: © 2021 Massachusetts Institute of Technology Published under a Creative Commons Attribution 4.0 International (CC BY 4.0) license

(Kuhl, 2014; Werker et al., 2012). This perceptual narrowing has a neurophysiological basis. Kuhl postulated that exposure to a particular language results in the development of dedicated (committed) neural networks which code the patterns of the language in question, and also make the learning of other languages more difficult (Kuhl, 2004). This native language neural commitment (NLNC) hypothesis has been influential in shaping our current understanding of how the nervous system supports perceptual narrowing and early language acquisition.

While the NLNC is able to explain a variety of behavioral and neurophysiological (in particular, cortical) data, its current rendition does not account for two important facts about speech sound processing in the developing nervous system. First, although infants in the first six months of life are able to perceive a large inventory of sounds, they are unable to perceive *all* sounds. In fact, it seems to be the case that infants, regardless of their language environment, have difficulty in discriminating speech sounds that are less acoustically salient (Best et al., 1995, 2016). In order for infants to distinguish whether or not a speech sound belongs to its native language and to use it contrastively in lexical contexts, they must first encode and recognize it. Second, speech sound development (and language acquisition) depends on a nervous system that is also developing, regardless of the specific language input, and this helps to determine how well infants can encode speech. Therefore, a comprehensive theory about early language acquisition must be situated in the context of neural development as well as language experience. The aim of the present study is to examine the development of speech perception in the context of both cortical and subcortical neural development during the first two years of life. Our main question is whether perceptual attunement can be explained by developmental changes in neural functions.

The auditory neural system that supports hearing and spoken language consists of both subcortical (including the auditory brain stem) and cortical structures that develop at different rates. In terms of anatomical growth, the connections in the auditory brain stem acquire their mature pattern by the 29th week of gestation and their myelination is completed within the first postnatal year (Hartley et al., 2010; Moore & Linthicum, 2001, 2007; Moore et al., 1995, 1997; Romand & Romand, 1982; Romand et al., 1976; Sano et al., 2007). The 6th postnatal month sees the maturation of the brain stem dendritic trees (Moore et al., 1998). The subcortical-cortical connections, on the other hand, only mature when children are between 3 and 5 years old, and the myelination of cortical layers continues up to the age of 11 or 12 (Moore & Linthicum, 2007).

In terms of functional growth, research on the human auditory neural systems has traditionally relied on scalp-recorded electroencephalography (EEG). Cortical and subcortical EEG components can traditionally be separated by their timing and frequency composition. High-frequency *early-latency* frequency-following response (FFR) is phase-locked to the spectral envelope of an auditory stimulus (Moushegian et al., 1973). Although much more research is needed to understand the sources of the FFR at different frequency ranges (Coffey et al., 2019), the available evidence has suggested that the FFR has stronger (though not exclusive) subcortical generators, including the inferior colliculi, especially for frequencies above 150 Hz (Bidelman, 2018; Tichko & Skoe, 2017). The field seems to be more certain that the auditory *long-latency* response (LLR) is generated exclusively in the cortex, within superior temporal gyri in the vicinity of the Heschl's gyrus (Godey et al., 2001) where intracranial recordings have shown the presence of phoneme categorical representations (Chang et al., 2010; Mesgarani et al., 2014). The adult LLRs in the time range between 50 and 240 ms are sensitive to the nativeness of phonemes in both active discrimination and passive listening (Gansonne

Frequency-following response (FFR): Scalp recordings of human phase-locked neural activity that are synchronized to periodic and transient aspects of sound.

et al., 2018; Kuuluvainen et al., 2014; Näätänen et al., 1997; Ross & Tremblay, 2009; Shtyrov et al., 2000; Tremblay & Kraus, 2002; Tremblay et al., 1998, 2001, 2014).

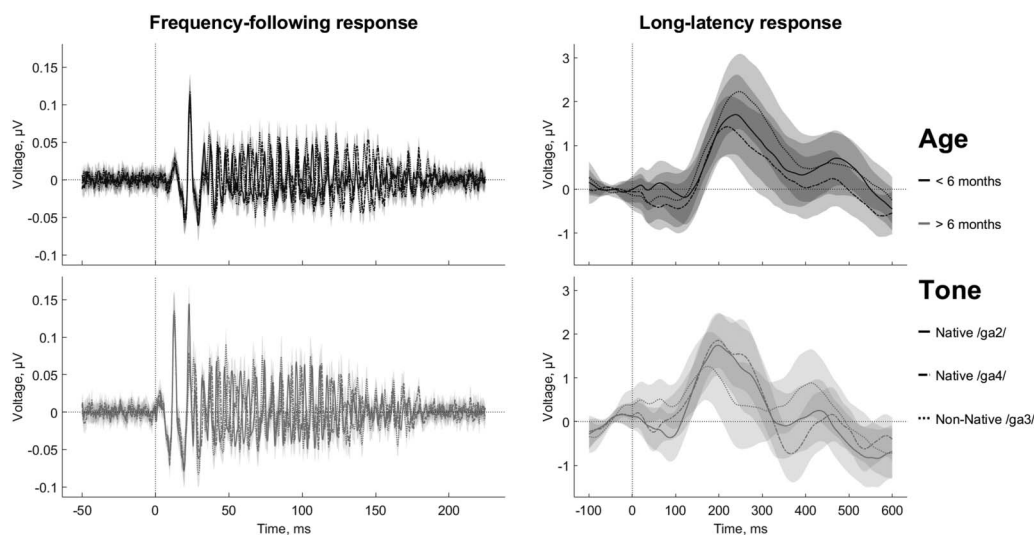
The FFRs to speech are clearly detectable at birth (Ribas-Prats et al., 2019). The time and frequency-domain parameters of FFR to consonants and vowels improve steadily from birth to 11 years (Anderson et al., 2012, 2015; Skoe et al., 2015). Neonates display a much less matured FFR to lexical tones, pitch patterns that are used to contrast word meaning, than do adults (Jeng et al., 2011, 2016). Although not much is known about the development of FFR to lexical tones in the first few years of life beyond the neonatal period, quite a few research studies have examined FFR to speech more generally in adulthood. Lifelong experience of lexical tones results in enhanced FFR encoding to lexical tones in tone language speakers relative to non-tone language speakers (Krishnan et al., 2005, 2009). Experience with other tone languages, e.g., Thai, may benefit the neural representation of lexical tones of another tone language (e.g., Mandarin; Krishnan et al., 2010). With short-term training, FFR to lexical tones can be improved in native speakers of a non-tone language (Reetzke et al., 2018; Song et al., 2008). Experience-dependent FFR responses may be partially mediated by cortico-fugal projections to the auditory brain stem (Intartaglia et al., 2016; Krishnan et al., 2005; Suga, 2008; Suga et al., 2000; Zhao & Kuhl, 2018) but may also manifest an online adaptation to ongoing statistics in the brain stem (Alho et al., 2019; Escera, 2017; Slabu et al., 2012). These results clearly demonstrate that the FFRs may be a useful tool for studying the neural development of speech. As far as we are aware, studies have yet to investigate the early development of FFR to native and non-native sounds during the period of rapid speech development.

Relative to FFR, a larger body of research has examined the LLR in relation to early speech development. LLR at term birth is dominated by a wide and late positive deflection peaking in the 200–300 ms interval (Barnet & Lodge, 1967; Barnet et al., 1975; Kushnerenko et al., 2002; Novak et al., 1989; Wunderlich & Cone-Wesson, 2006). Later in life the original positivity splits into P1 and P2, the trough between the positive peaks deepens and eventually by 12–16 years old, it becomes the N1 wave of the adult auditory event-related potential (ERP) (Cone & Whitaker, 2013; Lippé et al., 2009; Novak et al., 1989; Ponton et al., 2002; Shafer et al., 2015; Small et al., 2018). (Throughout this article, we use the term *P1* for the largest infant positivity and consider it the homologue of the whole P1-N1-P2 complex in the adult LLRs). More specific to speech, Kuhl and colleagues (Kuhl, 2010) conducted a series of studies to examine the cortical development of native and non-native speech perception. They employed mismatch negativity (MMN), a pre-attentive component of the LLR, that serves as an electrophysiological index of sound discrimination (Cheour-Luhtanen et al., 1995; Partanen et al., 2013). The MMN to various vowel contrasts is present in newborns (Cheour-Luhtanen et al., 1995), but it reduces in response to non-native vowel contrast after 6 months of age (Cheour et al., 1998; Ortiz-Mantilla et al., 2013). Neural commitment to the native language at the age of 7.5 months can predict language development up to two years following initial neural measurements (Kuhl et al., 2008; Rivera-Gaxiola, Klarman, et al., 2005; Rivera-Gaxiola, Silva-Pereyra, & Kuhl, 2005). A recent study of acoustic change complex in the infant LLR demonstrated a transformation from a merely acoustical-driven pattern to native phoneme-category sensitivity around the age of 6 months (McCarthy et al., 2019). Though the findings from these pioneering LLR-based (mostly MMN) studies provide an initial neural explanation for perceptual narrowing, these studies are focused largely on the role of the cortex. In order to obtain a more comprehensive understanding of the development of speech perception in the context of neural development, further research is needed that examines both cortical and subcortical levels of the auditory nervous system.

### The Current Study

In the present study, we examined the functional development of neural responses with generators in the brain stem and cortex in relation to native and non-native speech sounds. As the anatomical and neurophysiological development of the brain stem generally occurs earlier than that of the cortex, we postulate that the function of the brain stem in the first years of life is solely to increase the encoding accuracy of the acoustic input, regardless of its nativeness status. Thus, we predict a beneficial effect of age on the early-latency FFR with no interaction between age and nativeness status during the first two years of life. An increase in encoding fidelity would then allow the cortex to differentiate between native and non-native sounds, enabling greater neural resources to be devoted to native sounds to facilitate infants' phonological and lexical development. By contrast, long-latency cortical LLR to speech develop more slowly, and the responses show sensitivity to the nativeness status of the speech stimulus. We expect that the already known cortical sensitivity to the nativeness status of a language extends also to the P1 component of the infant LLR as they reflect interacting cortical processes.

Early-latency neural encoding FFR and long-latency cortical LLR recordings were simultaneously collected via EEG (Figure 1) from 85 Cantonese-learning children as they heard two native (Cantonese) and one non-native (Mandarin) speech stimuli that were different only in their pitch (Supplementary Figure 1; Supporting Information (SI) can be found at [https://doi.org/10.1162/nol\\_a\\_00049](https://doi.org/10.1162/nol_a_00049)). Both Cantonese and Mandarin are tone languages. Lexical tones can be ideal for testing our hypothesis because phase-locking to fundamental frequencies that represent lexical tones can be measured by FFR. Children's EEG responses were collected longitudinally at up to 4 time points (Supplementary Figure 2). Since not every child had a recording at every time point, the final data set represented a mixture of longitudinal trajectories and cross-sectional data points. We employed mixed-effects models that took account of repeated



**Figure 1.** The waveforms of the early-latency and long-latency responses differ in their timing, temporal structures, and maturational patterns. The *left panel* shows the frequency-following response generated predominantly in the brain stem. The *right panel* shows the waveforms of the long-latency responses generated in the cortex. Note the difference in time and amplitude scales between the two panels. The rows correspond to the age groups younger and older than 6 months that are also coded with darker and lighter shades of grey correspondingly. The three lines represent each of the three lexical tones: Native /ga2/ (solid), Native /ga4/ (dash-dotted), and Non-Native /ga3/ (dotted).

measures when the data were available within the same subjects. This design is similar to other studies of neural development (Raznahan et al., 2014; Wierenga et al., 2014).

We expected measures of FFR to improve linearly and substantially across age during the early stages of development (23 days to 24; 16 months), as predicted by the literature on brain-stem development (Moore et al., 1998). Importantly, improvement in these FFR measures can be observed in both native and non-native tones. By contrast, LLR measures are likely to improve more slowly than those of early-latency neural encoding measures within the same developmental period because the auditory cortex takes longer to develop (Moore & Linthicum, 2007). As in previous studies of speech development in the cortex, we expected to see sensitivity to the nativeness status of speech sounds occurring at about 6 months (Ortiz-Mantilla et al., 2013), because 6 months is when cortical responses to lexical tones are found to be adultlike (Cheng et al., 2013). As lexical tone is a type of speech sound that occurs in every syllable in a tone language and about 70% of the languages in the world are tone languages (Fromkin, 2014; Yip, 2002), we believe an investigation into the cortical-subcortical development of speech using lexical tone as a starting point is a reasonable one. Our results can lay the foundation for future research that examines other categories of speech sounds.

## MATERIALS AND METHODS

### Participants

Eighty-five Cantonese-learning infants and toddlers (48 males) between 23 days and 24; 16 months of age (mean age 7; 5 months, *SD* 5; 9 months), with no significant medical history, participated in this study. Most of the participants were recorded longitudinally with up to 4 visits 5; 7 months  $\pm$  7.2 days apart on an average, producing 177 data points (Supplementary Figure 2). By self report, parents were native Cantonese speakers and spoke Cantonese at home. Note that although English and Mandarin (Putonghua) are languages spoken by residents of and visitors to Hong Kong and are taught in the public schools, the vast majority of the population speaks Cantonese (88.9%) as their usual language, with English (4.3%) and Mandarin (1.9%) being the clear minorities (Government of Hong Kong SAR, 2016).

Written informed consent was obtained from parents or guardians prior to their child's participation in the study. The experimental protocol was in compliance with the Declaration of Helsinki.

### Stimuli and Experimental Design

The lexical tones were Cantonese (Native) tones 2 and 4 (rising and falling) and Mandarin (Non-Native) tone 3 (dipping) embedded in the syllable /ga/. We chose the speech stimuli based on two principles: (1) The native tone categories must be distinctive; tone 2 and tone 4 being of different contour shapes fulfilled this requirement. (2) The native and non-native tone categories must also be distinctive. Mandarin tone 3 is quite unique and does not overlap with any Cantonese tones. Furthermore, these same speech stimuli were used in our previous study (Maggu et al., 2018), attesting to their validity for obtaining robust FFR responses.

The stimuli were spoken by a phonetically-trained female speaker, and were duration and amplitude normalized. The fundamental frequency ( $F_0$ ) ranges for the Native /ga2/, Native /ga4/, and Non-Native /ga3/ were respectively 182–278 Hz, 152–187 Hz, and 142–177 Hz (Supplementary Figure 1). These three tones are described as 25, 21, and 214, respectively, according to Chao's nomenclature (Chao, 1930). Those frequencies are well above the frequency range for cortical generators to be the dominant sources (Bidelman et al., 2018).



Non-native tone was acoustically more complex than the other tones (see SI). This acoustic complexity was not a confound because our results suggested that the Native /ga2/ tone posed the most difficulty for the listeners (see SI). The stimuli were presented to the participants via Audio CPT module of STIM2 (Compumedics, Australia). One thousand repetitions of the tones were delivered in separate blocks to both ears of the participant with fixed interstimulus interval 500 ms (stimulus-onset-asynchrony 674 ms) and alternating polarity.

### **EEG Recording and Analysis**

Continuous EEG was collected using a Synamps RT amplifier connected to a Curry 7.05 workstation (Compumedics, Australia) from five Ag/AgCl electrodes at Cz, M1 (left mastoid) and M2 (right mastoid) at a sampling rate of 20 kHz with CPz as reference and Fpz as ground. Interelectrode impedances were maintained at less than or equal to 1 k $\Omega$ . Cz data were re-referenced offline with the average of two mastoids. The subsequent analysis was performed on Cz. All data analysis was performed in MATLAB 9.2 (R2017a; [mathworks.com](http://mathworks.com)), with custom-made scripts using built-in functions of EEGLab (Delorme & Makeig, 2004), ERPLAB (Lopez-Calderon & Luck, 2014), Signal Processing Toolbox ([mathworks.com](http://mathworks.com)), and Statistics and Machine Learning Toolbox ([mathworks.com](http://mathworks.com)). Two different analysis pipelines were used for FFR and LLR extraction correspondingly (see below). For each of the early-latency neural encoding FFR and long-latency cortical LLR responses, we examined a number of conventional metrics to investigate their developmental trajectories. An important consideration was whether the native and non-native tones differ in their developmental trajectories of the auditory early-latency neural encoding and the long-latency cortical responses.

Our recording procedures allowed us to simultaneously collect early-latency FFR and long-latency LLR with the same set of stimuli in the same child. As far as we are aware, this is the first study to simultaneously record FFR and LLR in infants; only adult studies of the kind have been performed to date (Alho et al., 2019; Shiga et al., 2015). For each of the early-latency FFR and long-latency cortical LLR measures, we examined a number of conventional metrics (Anderson et al., 2012, 2015; Jeng et al., 2011, 2013, 2016; Skoe et al., 2015) to investigate their developmental trajectories. The aim was to trace all the subjects longitudinally. However, the main challenge of a longitudinal design was subject attrition, which turned this study into a mixed cross-sectional and longitudinal design. Additionally, the EEG recording in infants is prone to movement artifacts and not every recording session is successful. One option would be keeping only the subjects with a certain amount of longitudinal data points. The size of our sample would allow us such a manipulation. However that would not only reduce statistical power but also render the sample unrepresentative in relation to the original population. We opted to keep all subjects who had at least one successfully recorded EEG session. To take advantage of the fact that repeated measurements were taken from some participants, we employed mixed-effects models with subject as a factor. Our primary focus was to understand whether the growth of FFR measures was rapid, and whether this growth could be observed in both native and non-native tones. Our secondary focus was to understand whether the growth of the LLR measures appeared to be less substantial, and whether sensitivity to native-ness could be seen at around 6 months of age.

### **FFR Analysis**

For FFR extraction, the data were filtered with an 8-order Butterworth 80–1500 Hz bandpass filter and down-sampled to 3 kHz. The epochs were extracted around the stimulus onset time

from –50 to 225 ms using a  $\pm 25$   $\mu\text{V}$  artifact rejection criterion (Skoe & Kraus, 2010). The EEG recording blocks with more than 10% of the rejected sweeps per block (i.e., >100 rejections) were not included in further analyses (Maggu et al., 2016, 2018). Thus, the minimum number of trials per block was 900. One block comprised the EEG responses to one tone-type for a particular subject at a particular age. The averaged FFR with more than 0.7  $\mu\text{V}$  maximum were excluded for suspicion of stimulus artifact based on preliminary data screening. These procedures resulted in exclusion of 18% of the EEG blocks, including the whole set of data from one participant. In addition, 13% Native /ga2/ blocks, 7% Native /ga4/ blocks, and 9% of the Non-Native /ga4/ blocks were missing in the final data set. The final number of trials per block was  $988.8 \pm 1.5$  for Native /ga2/,  $990.0 \pm 5.17$  for Native /ga4/, and  $984.8 \pm 2.0$  for Non-Native /ga3/. Neither the difference between the tones for the number of accepted trials nor the correlation between the number of trials with age was significant (Native /ga2/ vs. Native /ga4/:  $t(316) = -0.243$ ,  $p = 0.808$ ; Native /ga2/ vs. Non-Native /ga3/:  $t(313) = 1.597$ ,  $p = 0.111$ ; Native /ga4/ vs. Non-Native /ga3/:  $t(323) = 0.981$ ,  $p = 0.327$ ; Native /ga2/ vs. age:  $r = 0.15$ ,  $p = 0.063$ ; Native /ga4/ vs. age:  $r = 0.097$ ,  $p = 0.215$ ; Non-Native /ga3/ vs. age:  $r = 0.054$ ,  $p = 0.5$ ).

FFR was calculated as a time-domain average of all good epochs with baseline correction across the pre-stimulus interval (–50 to 0 ms).

FFR pitch contour was extracted with a sliding time window applied at the time-frequency decomposition of intertrial phase coherence (a.k.a., phase-locking factor, Tallon-Baudry et al., 1996). In comparing FFR pitch contour with auditory stimulus pitch contour, the latter was shifted 20 ms toward the end of the time course for the pitch tracking accuracy calculation (see Jeng et al., 2013 and SI for details).

A number of time-domain and frequency-domain metrics were extracted from FFR. For the main report on FFR here, we selected the four measures most commonly employed in FFR developmental studies: lower-frequency and middle-frequency spectral band-power (Anderson et al., 2012, 2015; Skoe et al., 2015), pitch strength (Jeng et al., 2011, 2013, 2016) and signal-to-noise ratio (SNR). Kolmogorov-Smirnov tests confirmed the normality of the four measures in question (see SI). The FFR SNR was calculated as a dB-transformed ratio between root mean square power of post-stimulus and pre-stimulus intervals of FFR (see SI for the formula). Whole-epoch autocorrelation was calculated on FFR waveforms. Pitch strength was calculated as peak-to-peak distance on an autocorrelation curve between the first autocorrelation peak and the preceding trough. Lower and middle-frequency spectral band-powers were obtained by calculating the fast Fourier transform of FFR waveforms and extracting the power around the F0 and its two lowest overtones (2D and 3D harmonics). See SI and Supplementary Figure 3 for the description of the other metrics.

The aforementioned results focused mostly on the encoding of individual tones. One important requirement of phonological development is to differentiate distinct speech sound categories. We therefore examined classification of the three tones with a support vector machine (SVM) for each month of age that delivered more than 10 observations, i.e., from birth to 14 months. Averaged FFRs were down-sampled to 120 Hz in order to reduce the number of features in the models. The interval from 0 to 150 ms was extracted, resulting in 13 time points/features per observation. The data at each time point were normalized across observations. The data of each age were bootstrapped 10,000 times. In each bootstrap sample, we ran SVM classification with cross-validated accuracy as the outcome. The procedure was performed for both real tone labels and randomly permuted labels labels. (See SI for the detailed SVM parameters.)

### LLR Analysis

For LLR extraction, the data were filtered with a 4-order Butterworth 0.1–30 Hz bandpass filter and down-sampled to 3 kHz. The epochs were extracted around the stimulus latency time from –100 to 600 ms using a  $\pm 100$   $\mu$ V artifact rejection criterion. All epochs with amplitude in the artifact rejection criterion were considered good epochs. Since LLR is known to habituate with the repetition of the same sound (Cortesa et al., 2019; Näätänen & Picton, 1987), only 300 first good epochs for each participant/condition were used for the purposes of analysis. LLR was calculated as a time-domain average with baseline correction across the pre-stimulus interval (–100 to 0 ms). SNR was calculated on an averaged LLR with a pre-stimulus interval from –100 to 0 ms and a post-stimulus interval from 0 to 600 ms (see SI for the formula). The LLR main positive peak (i.e., P1) was searched for automatically as a maximum across the post-stimulus interval (0 to 600 ms; see SI for details). Similarly to the FFR analyses, we employed linear mixed-effects (LME) models with age and tone as fixed effects and participant as a random intercept.

Previous research on the cortical development of lexical tones has demonstrated that 6 months of age is an important milestone for adultlike responses, as the polarity of MMN switches from a positive to a negative (adultlike) response during this period (Cheng et al., 2013). This finding converges with the studies of vowels and consonants that have shown a decrease in MMN responses to non-native sounds in infants older than 6 months of age (Cheour et al., 1998; Rivera-Gaxiola, Silva-Pereyra, & Kuhl, 2005). We tested cortical development of lexical tones by examining age-dependent effects categorically, using 6 months (165 days) as a cutoff. We took a step further to analyze the effects of nativeness by categorizing our stimulus into native (pooled Native /ga2/ and Native /ga4/) and non-native (Non-Native /ga3/) tones with a focus on the latency and amplitude measures. Because P1 peak latency and amplitude did not meet the normality assumption, we used non-parametric tests to investigate the hypothesis of age-dependent sensitivity to speech nativeness.

### FFR vs. LLR Comparison

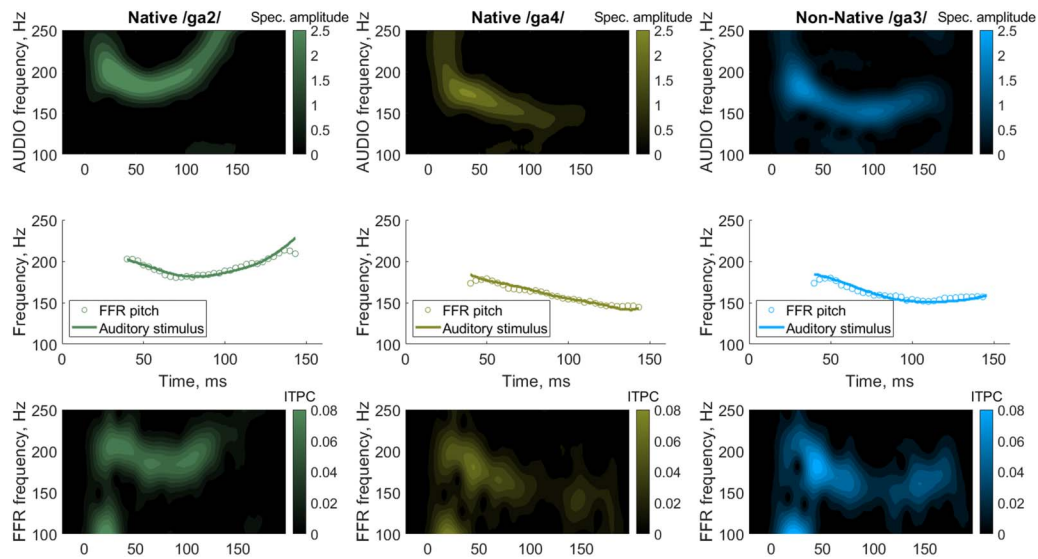
In order to compare the magnitude of growth between the early-latency neural encoding FFR and long-latency cortical LLR measures, we examined three measures that can be obtained in both FFR and LLR, namely, SNR, amplitude, and latency. We examined a group of younger ( $n = 22$ ) and older ( $n = 17$ ) children with an arbitrary age cutoff of younger than 42 days and older than 400 days, as that would provide the maximum number of participants in each age group. Extracting the effect size of age cutoff (i.e., the slope of the regression line between the two age groups) on the normalized metrics enabled us to compare the maturation of the metrics directly, regardless of individual means and variance, where maturation was indicated by larger SNR, increased amplitude, and shorter latency.

## RESULTS

### Early-Latency Neural Encoding FFR Measures

The FFR consists of a series of transient peaks followed by a steady state response (Figure 1, left panel). We first examined whether, on average, the infants' FFRs showed high fidelity. For each tone, we calculated the intertrial phase coherence extracted pitch contour, which followed the pitch contour of the stimuli when the time lag equal to the averaged latency of stimulus-response correlation (20 ms) was introduced. We found that the averaged pitch contours were strongly correlated with the pitch contours of the speech stimuli, with  $r^2$





**Figure 2.** The pitch contour of the frequency-following response (FFR) follows the pitch contour of the auditory stimuli. The columns correspond to the lexical tones Native /ga2/, Native /ga4/, and Non-Native /ga3/. *Upper row:* spectrogram of the auditory stimuli. *Middle row:* the overlap of the auditory (black line) and early-latency brain-response (grey circles) pitch contours. The pitch tracking (audio vs. brain signal correlation) is 0.91 for Native /ga2/, 0.97 for Native /ga4/, and 0.95 for Non-Native /ga3/. *Bottom row:* Grand-averaged time-frequency decomposition of the intertrial phase coherence.

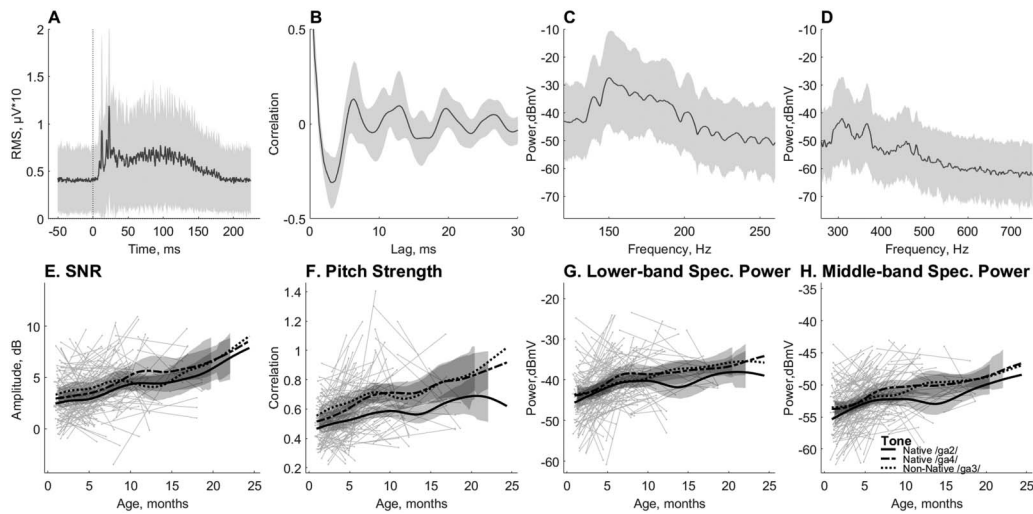
values of 0.91 for Native /ga2/, 0.97 for Native /ga4/, and 0.95 for Non-Native tones (Figure 2). This suggests that as a group, our infant participants were quite accurate in encoding the speech stimuli.

**LME analysis**

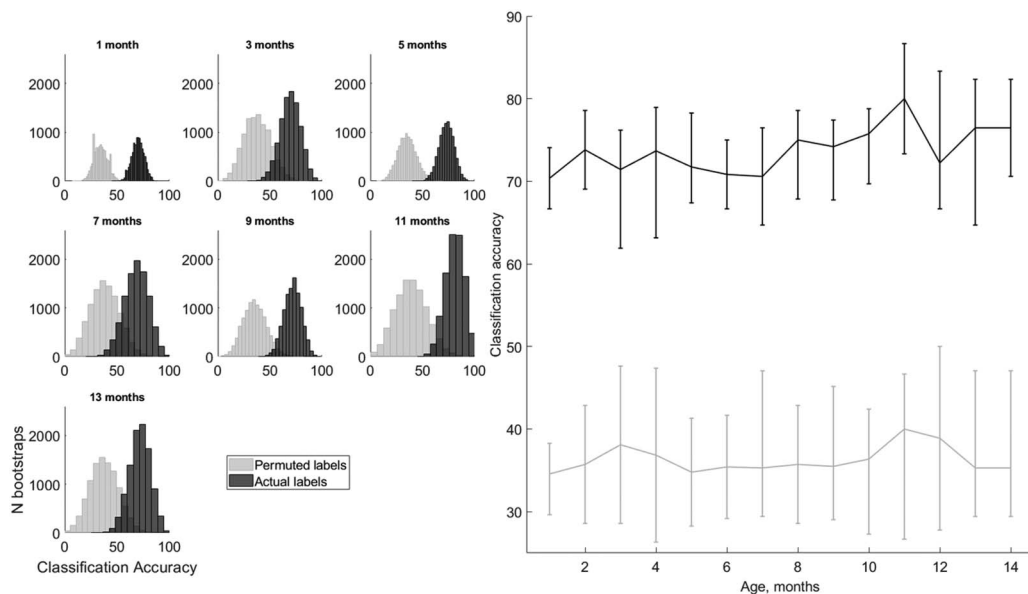
*F* tests of the fixed effects within the mixed-effects model revealed a main effect of age for FFR SNR ( $F(1, 473) = 19.21; p = 1.44 \cdot 10^{-5}; pBf = 0.0002; \eta_p^2 = 0.04$ ), lower-frequency spectral band-power ( $F(1, 473) = 10.91; p = 0.0010; pBf = 0.0133; \eta_p^2 = 0.02$ ), and middle-frequency spectral band-power ( $F(1, 473) = 17.48; p = 3.47 \cdot 10^{-5}; pBf = 0.0005; \eta_p^2 = 0.04$ ). The main effect of age on pitch strength was marginally significant ( $F(1, 473) = 7.24; p = 0.0074; pBf = 0.0958; \eta_p^2 = 0.02$ ). There was a marginal effect of tone for pitch strength only ( $F(2, 473) = 4.12; p = 0.0169; pBf = 0.2198; \eta_p^2 = 0.02$ ), while it was not significant for FFR SNR, lower-frequency spectral band-power and middle-frequency spectral band-power (all  $p > 0.05$ ). Most importantly, none of the FFR measures revealed an interaction between age and tone, including FFR SNR, lower-frequency spectral power, middle-frequency spectral power, and pitch strength (all  $p > 0.05$ , Figure 3).

**Classification analysis**

We found that SVM classification of the early-latency brain responses to the three tones improved over time (Figure 4). The median classification of the tone was overall higher than the permuted classification accuracy (mean  $73.8 \pm 0.75\%$  and  $36.3 \pm 0.43\%$ ,  $t(13) = 53.48; p = 1.25 \cdot 10^{-16}; d = 14.3$ ). Importantly, classification accuracy improved with age as indicated by the Pearson correlation ( $r = 0.64; p = 0.0136; pBfr = 0.0272; d = 1.67$ ). Adding a quadratic term did not improve the model, suggesting a linear dependency ( $F(1, 11) = 0.14; p = 0.71; \eta_p^2 = 0.01$ ). No age dependency was found for the permuted data classification ( $r = 0.24; p = 0.407; pBfr = 0.814; d = 0.49$ ).



**Figure 3.** Developmental trajectories of early-latency brain measures show linear growth with age, regardless of speech nativeness. The top panels show the root mean square (RMS) of the frequency-following response (FFR) (A); the FFR autocorrelagram (B); and the lower (120–260 Hz) (C) and middle (260–750 Hz) (D) frequency bands of the FFR spectrum. The shaded area marks the standard deviation. The bottom panels plot the output metrics as a function of age in days. The metrics are the signal-to-noise ratio (SNR) of the time-domain waveform (E); the pitch strength (trough-to-peak magnitude of the autocorrelagram) (F); and the mean power in the lower (G) and middle (H) frequency bands. Individual data points are plotted as grey dots with lines connecting the data of the same participant and the same tone. The three lines represent the average age-dependent development for each of the three lexical tones Native /ga2/ (solid), Native /ga4/ (dash-dotted), and Non-Native /ga3/ (dotted). The data are smoothed with Eilers’ technique (Eilers, 2003).

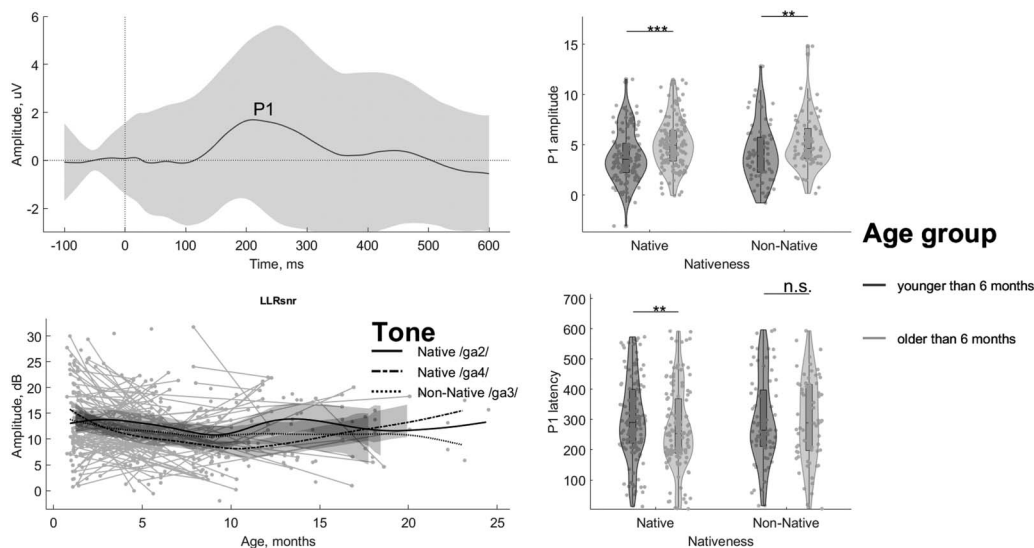


**Figure 4.** Tone classification with support vector machine improves with age for actual but not for permuted tone labels. The left panel shows real (black) and permuted (grey) labels for the cross-validated classification accuracy distribution across 10,000 bootstrapped samples for selected ages in 2-month steps from 1 to 13 months. The right panel shows the medians of those distributions as a continuous function of age in months from 1 to 14 for real (black) and permuted (grey) labels. The error bars show the range between the first and third quartile of the distributions. The real label classification accuracy correlates significantly with age ( $r = 0.64$ ,  $p < 0.05$ ,  $d = 1.67$ ), while the permuted label classification does not ( $r = 0.24$ ,  $p > 0.1$ ,  $d = 0.49$ ).

### Long-Latency Cortical LLR Measures

The LLR was dominated by a large and wide positive peak P1 (Figure 1, right panel). We measured the latency and amplitude of P1, as well as its SNR. As in the case of FFR, we investigated the changes of LLR measures as a function of age (Figure 5). In the LME results, there was an increase with age in P1 amplitude ( $F(1, 409) = 9.14; p = 0.0027; pBf = 0.0345; \eta_p^2 = 0.02$ ), but not in LLR SNR ( $p > 0.05$ ) or P1 latency ( $p > 0.05$ ). None of the LLR parameters were different between tones, i.e., LLR SNR, P1 amplitude, or P1 latency (all  $p > 0.05$ ). There was no interaction between age and tones for any of the LLR parameters either, i.e., LLR SNR, P1 amplitude, and P1 latency (all  $p > 0.05$ ). In general, the effect of age on LLR is much less robust compared with the FFR results.

We then followed the findings of the literature concerning the cortical development of lexical tones at 6 months of age (Cheng et al., 2013). For non-native speech, the age split resulted in 63 participants (71 data points) and 51 participants (68 data points), respectively, in the younger and older groups in the analysis. For native speech, 65 participants (135 data points for pooled tones, 75 data points for averaged tones) and 61 participants (141 pooled tone data points, 89 averaged tone data points) were in the younger and older groups, respectively. The Wilcoxon rank-sum tests demonstrated a significant shortening of P1 latency and an increase in its amplitude after the age of 6 months for the native tones (pooled Native /ga2/ and Native /ga4/:  $Z = -2.45, p = 0.014, r = -0.15$  for latency and  $Z = 4.04, p = 5.27 \cdot 10^{-5}, r = 0.24$  for amplitude; averaged Native /ga2/ and Native /ga4/:  $Z = -2.41, p = 0.016, r = -0.19$  for latency and  $Z = 3.87, p = 1.07 \cdot 10^{-4}, r = 0.30$  for amplitude), while the change was relatively small for the Non-Native /ga3/ tone ( $Z = 0.16, p = 0.56, r = 0.01$  for latency and  $Z = 2.52, p = 0.006,$



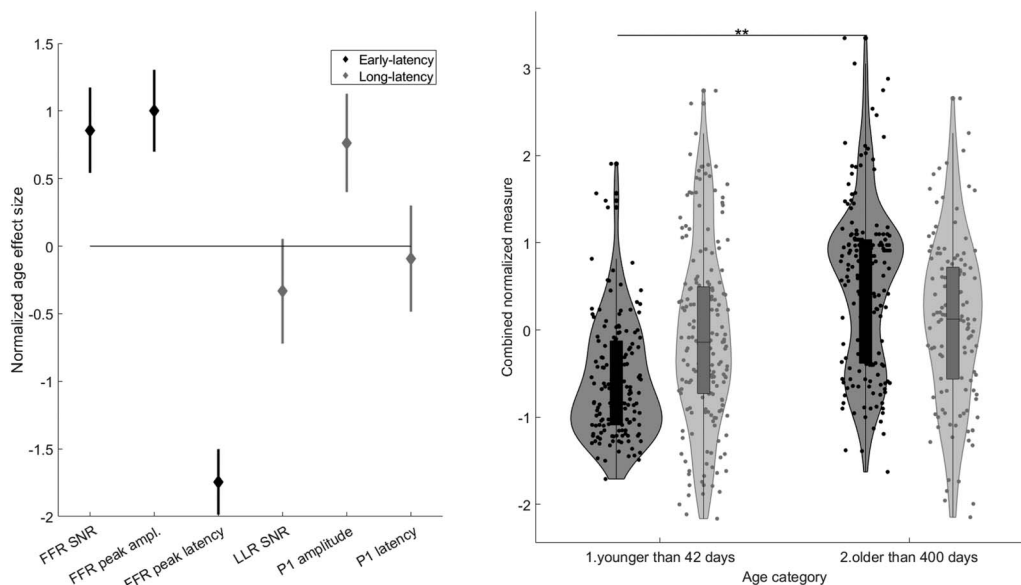
**Figure 5.** The long-latency response (LLR) signal-to-noise ratio (SNR) does not grow with age, while its peak latency reduces only for the native tones. Upper left panel shows the LLR time-domain waveform averaged across all data points, with the shaded area marking the standard deviation. Peak P1 is labeled at maximum. Lower left panel shows LLR SNR as a function of age in days. Individual data points are plotted as grey dots with lines connecting the data of the same participant and the same tone. The three lines represent the average age-dependent development for each of the three lexical tones Native /ga2/ (solid), Native /ga4/ (dash-dotted), and Non-Native /ga3/ (dotted). The data are smoothed with Eilers' technique (Eilers, 2003). Upper and lower right panels show violin plots of P1 amplitude and latency distributions across the 6-month cutoff. The tones are pooled into native (/ga2/ and /ga4/, dark-grey) and non-native (/ga3/, light-grey) groups. The grey circles represent individual data and the diamonds mark the medians of the distribution. The significance of the age effect in the Wilcoxon rank-sum test is illustrated with asterisks separately for native and non-native tones (\*\* $p < 0.01$ , \*\*\* $p < 0.001$ , uncorrected, n.s. non-significant).

$r = 0.21$  for amplitude) (Figure 5, right panel). Overall, these results replicated earlier findings that perceptual narrowing is supported by cortical processes (Conboy & Kuhl, 2011). Specifically, our results here confirmed that sensitivity to the nativeness status of lexical tones occurs at about 6 months of age, as suggested by a previous study (Cheng et al., 2013).

### Early-Latency Neural Encoding FFR vs. Long-Latency Cortical LLR Measures

As shown in the left panel of Figure 6, the effect size of FFR SNR (0.86, 95% CI [0.54, 1.17]) was significantly larger than that of LLR SNR (−0.33, 95% CI [−0.72, 0.05]). The effect size of FFR peak amplitude (1.00, 95% CI [0.70, 1.30]) was marginally larger than that of LLR P1 amplitude (0.76, 95% CI [0.40, 1.13]). Finally, the effect size of FFR peak latency was significantly larger (i.e., more negative, −1.74, 95% CI [−1.99, −1.50]) than that of LLR P1 latency (−0.09, 95% CI [−0.48, 0.30]). In other words, two out of the three cortical measures showed less change over time, and two out of the three measures showed a significantly larger age-related improvement in the early-latency neural encoding responses.

In order to simultaneously compare the effects of age category on the multiple parameters of early- and long-latency responses, we entered the aforementioned normalized metrics into a combined LME analysis with Age category, Tone, ERP type (early-latency vs. long-latency responses), and ERP measure (SNR vs. latency vs. amplitude) as fixed effects and Subject as random intercept.  $F$  tests of the fixed effects demonstrated the significant main effect of Age category ( $F(1, 620) = 9.92, p = 0.0017; \eta_p^2 = 0.02$ ) and the interaction between Age category and ERP type ( $F(1, 620) = 5.62, p = 0.0174; \eta_p^2 = 0.01$ ). None of the other main effects or interactions was significant (see Supplementary Table 3). In the follow-up split-by-ERP-type LMEs, the metrics of the early-latency ( $F(1, 317) = 10.6, p = 0.0013; \eta_p^2 = 0.03$ ), but not of



**Figure 6.** The maturational effect is greater for early-latency (black) than long-latency (grey) response measures. *Left panel:* Both brain subdivisions are represented by three compatible time-domain measures; signal-to-noise ratio (SNR), peak amplitude, and peak latency. The diamonds represent effect sizes measured as the difference between the means of the youngest (younger than 42 days) and oldest (older than 400 days), with participants’ data normalized for each measure. Lines cover the confidence intervals of those effect sizes in the linear models. The confidence intervals for SNR and peak latency do not overlap between the frequency-following response (FFR) and the long-latency response (LLR), indicating significant differences. *Right Panel:* The linear mixed effects analysis shows that a combination of normalized measures grows with age for early-latency (black) but not for long-latency (grey) response measures (\*\* $p < 0.01$ ).

the long-latency responses ( $F(1, 303) = 0.018, p = 0.8930; \eta_p^2 < 0.001$ ), were improved by age (right panel of Figure 6).

## DISCUSSION

The goal of the present study was to investigate the contributions of early-latency neural encoding and long-latency cortical processing toward early language development in the first two years of life. We postulated that a degree of encoding accuracy of speech sounds regardless of native status must occur in the early stages of life, before these speech sounds' native status could be determined. Consistent with this hypothesis, we found that the early-latency neural encoding of speech grows linearly and markedly for both native and non-native tone categories for children who are learning a tone language. In contrast, the growth in the long-latency cortical responses is much slower, but crucially the change is sensitive to the nativeness status, with changes occurring at about 6 months of age, as reported in previous studies (Cheour et al., 1998; Cheng et al., 2013; Rivera-Gaxiola, Silva-Pereyra, & Kuhl, 2005). We argue that our results reflect the complementary roles of neural structures that are associated with earlier and later stages of processing along the ascending auditory neural pathway. For the frequencies that we examined, these neural structures are likely located in the brain stem and cortex for earlier and later stages of encoding and processing, respectively. While the brain stem (and potentially other neural structures) encodes speech sounds faithfully in an earlier stage of processing, the cortex assigns phonological information to them at a later stage. These two processes together are foundational to early language development.

Behavioral studies of lexical tones suggest that perception of lexical tones emerges at or before 6 months of age for tone-learning infants (Mattock & Burnham, 2006; Yeung et al., 2013). However, it is unclear how development of lexical tone perception is associated with early-latency and long-latency neural responses. As far as we are aware, the present study is the first to address this question. Our long-latency cortical findings are largely consistent with those of the studies of vowel and consonant perception (Cheour et al., 1998; Rivera-Gaxiola, Silva-Pereyra, & Kuhl, 2005) showing native and non-native differentiation in MMN responses after 6 months (and sometimes earlier). Although Cheng et al. (2013) did not investigate native and non-native tones, their findings of an adultlike cortical MMN response to native Mandarin tones in 6-month-old Mandarin-learning infants are generally consistent with the results in the present study. The MMN paradigm favored by the majority of previous LLR studies amplifies the brain responses to the phoneme categories by subtracting standards from deviants in oddball sequence (Escera, 2017; Näätänen & Winkler, 1999). However, in adults the sensitivity to phonological contrasts can be found already in the so-called obligatory brain responses (Grimaldi, 2018). The amplitude of responses to native phonemic contrast was found to increase with age in an equiprobable ERP paradigm (Key et al., 2012). We extended the existing MMN studies by investigating the difference between the obligatory P1 response in infant LLR to stimuli in all-standard blocks. The results converge with the MMN findings as we found a latency delay for non-native tone that is likely to be caused by difficulty in categorization after attunement to native phonology at around 6 months. It is clear that multiple cortical LLR components are responsive to the native status of speech sounds. We must acknowledge that the conclusions derived from our cortical results were based on conducting less robust statistical analyses separating the native and non-native data, rather than relying on findings from a significant Age  $\times$  Nativeness interaction. Notwithstanding that, the results converged with previous studies.

In terms of early-latency speech encoding processes, Jeng et al. (2013, 2016) were the first to study FFRs in tone-learning infants. These two smaller-scale studies found that neonates



from Mandarin-speaking families did not differ from neonates from English-speaking families in their FFRs to Mandarin tones. This suggests the cross-language differences in FFRs to lexical tones found in studies of adults (Krishnan et al., 2010) are most likely attributable to language experience. It should be noted, however, that the developmental time course of FFR to lexical tones and the potential within-subject differences between native and non-native lexical tone encoding have not been investigated before. Two other studies specifically focused on the development of the FFR signal in early years and across the lifespan by examining non-tone language-speaking listeners, but they did not examine lexical tones or the developmental difference between native and non-native speech perception (Anderson et al., 2015; Skoe et al., 2015).

Neurophysiological data suggest that the roles of the brain stem and cortex and the relationship between them may change considerably as the infant grows. Evolutionary older brain-stem structures take shape earlier and mature faster than the neocortex (Moore & Linthicum, 2007). Postmortem fetal studies show active processes of axon growth, dendrite tree arborization, and myelination in the brain stem, which start in the last prenatal month and are mostly completed by the end of the first postnatal year (Hartley & King, 2010; Moore et al., 1995, 1997, 1998; Moore & Linthicum, 2007; Sano et al., 2007). It is the neocortex that empowers humans with speech, but the development of cortical structure achieves its adult-like condition only by the age of 11–12 years (Moore & Guan, 2001). Indeed, the functioning of the auditory cortex in the perinatal period is physiologically very different from its functioning in adults in terms of the cortical layer domination (Moore & Linthicum, 2007). The immature status of the cortex implies also the immature condition of the corticofugal projections to the brain stem. Thus, the potential contribution of the corticofugal system to speech perception in adults may be related to its maturation (Suga et al., 2000). On the other hand, recent studies have clearly shown that the adult midbrain is not a mere relay and is capable of detecting regularities in the auditory stream (Malmierca et al., 2019). This capability also needs time to mature. Our data confirm previous descriptions using non-invasive EEG recordings. We have followed several non-invasive measures of early-latency neural encoding that are likely to reflect the spread of myelination and increase in synaptic activity in the brain stem and structures connected to it. Moreover, we show that cortical responses grow more slowly and that their maturation is also native-language specific. We anticipate that further studies will combine functional methods such as EEG with non-invasive structural methods such as diffusion MRI in order to obtain a better understanding of the neurophysiological basis of speech acquisition.

Neural phase-locking to the stimulus envelope is the most accurate representation of the physical properties of the stimulus. It requires accurate timing of neural firing, i.e., a strong degree of myelination. The capability of neurons in phase-locking to the high frequencies reduces along the auditory axis, and in the cortex the spatial coding of frequency (tonotopy) is prevalent instead (Leaver & Rauschecker, 2016). The FFR may correspond to the phase-locked activity of neurons of different neural centers across the ascending auditory pathway. The empirical studies available to date seem to suggest that phase locking at higher frequencies (>150 Hz) in the FFR most likely originates from the auditory brain stem (Bidelman, 2018; Tichko & Skoe, 2017). The basic conceptualization proposed here is not restricted to the brain stem or another particular neural structure per se, but rather suggests a division of labour of different auditory neural centers during language development. It seems likely that some centers contribute to representing basic acoustic elements of speech via direct phase-locking to the stimulus, while others assign function to those acoustic elements. They both need to be in place in order for spoken language to develop. It is worth emphasizing that our present findings do not

inform—nor do they depend on—our understanding of the precise neuroanatomical origins of the early-latency neural speech encoding. Our results merely suggest that not all neural responses to speech are equal in their neurodevelopmental trajectories, and most importantly the differences between temporally earlier FFR and later LLR are important for a child to succeed in acquiring native speech. Future research should examine the exact locations of the neural generators of these two types of responses.

We studied lexical tone and speakers of one tone language as the starting point to investigate the development of FFR and LLR in relation to native and non-native speech sounds. The broader applicability of these findings must be ascertained through future studies of other speech categories and young children learning other languages. Future research must also consider the impact of sleep on neural measurements. Although the vast majority of our participants were sleeping naturally during the experiment, it was not clear at which exact stage of the sleep cycle, and a few were even awake during the EEG recording. Research has demonstrated that different sleep stages may affect the amplitude of the LLR, specifically P3 at the latency of around 600–700 ms (Barnet et al., 1975). Although the focus of our LLR analysis was on the P1 component, we could not completely rule out the influence of sleep stage on our results. We expected that our large sample size and recordings at multiple time points would reduce the impact of sleep on our findings. Nonetheless, we must acknowledge that the lack of information about sleep stages from our participants might have affected the results. Therefore, an important future direction of this line of research is to establish to what extent our results are reproducible in the fully awake infants.

In summary, our findings suggest that neural centers at the earlier stages of the ascending auditory pathway are crucial for native language development because they provide an accurate representation of an incoming acoustic signal, regardless of whether or not those sounds are native. This accurate representation would allow neural centers at later stages of the auditory pathway to determine their linguistic relevance. Our study provides a critical piece of information about language development that, to the best of our knowledge, no previous study has considered. It opens up a new line of inquiry for future research. For example, future studies should examine the minimum required encoding accuracy in the brain stem and other structures before the signal is sufficiently accurate for linguistic processing in the cortex. They should investigate the age at which early-latency neural encoding contributes to language development when the child is older, and the extent of this contribution. Given that long-term sound exposure can modulate early-latency neural encoding in adulthood (Krishnan et al., 2005, 2010; Wong et al., 2007), future studies should examine the age at which the influence of auditory experience on neural encoding of native speech emerges. In terms of clinical application, an understanding of the division of labour among different centers of the auditory pathway would enable diagnostic protocols to be developed for investigating different subtypes of central auditory processing and language problems (Ingvalson & Wong, 2013; Ettliger et al., 2014). While some disorders may be a result of poor encoding at lower-level centers, others may stem from a failure to assign accurate phonological categories in the cortex.

#### **ACKNOWLEDGMENTS**

The authors wish to thank the families for participating in this study, many of whom were participants in the Stanley Ho Developmental Cohort Study. We would also like to thank Lydia Leung, Agnes Chou, Debby Ma, and Charlene Chiu for their assistance. This work was supported by the Dr. Stanley Ho Medical Development Foundation and a grant from the Innovation and Technology Fund (Hong Kong) #ITS/067/18. Some of the results reported in

this article were presented at the 2019 Annual Meeting of the Organization for Human Brain Mapping.

#### FUNDING INFORMATION

Patrick Wong, Innovation and Technology Commission - Hong Kong (<https://dx.doi.org/10.13039/501100007156>), Award ID: ITS/067/18. Patrick Wong, Dr. Stanley Ho Medical Development Foundation.

#### AUTHOR CONTRIBUTIONS

**Nikolay Novitskiy:** Data curation, Formal analysis, Software, Validation, Visualization, Writing – original draft. **Akshay R. Maggu:** Conceptualization, Methodology, Investigation, Writing – review & editing. **Ching Man Lai:** Data curation, Investigation, Project administration, Writing – review & editing. **Peggy H. Y. Chan:** Data curation, Investigation, Writing – review & editing. **Kay H. Y. Wong:** Data curation, Investigation, Writing – review & editing. **Hugh Simon Lam:** Supervision, Resources, Writing – review & editing. **Tak Yeung Leung:** Resources, Writing – review & editing. **Ting Fan Leung:** Supervision, Resources, Writing – review & editing. **Patrick C. M. Wong:** Supervision, Conceptualization, Methodology, Project administration, Funding acquisition, Resources, Writing – original draft.

#### COMPETING INTERESTS

Patrick C. M. Wong is the founder of a startup company supported by a Hong Kong SAR Government startup scheme for universities.

#### REFERENCES

- Alho, K., Żarnowiec, K., Gorina-Careta, N., & Escera, C. (2019). Phonological task enhances the frequency-following response to deviant task-irrelevant speech sounds. *Frontiers in Human Neuroscience, 13*, Article 245. <https://doi.org/10.3389/fnhum.2019.00245>, PubMed: 31379540
- Anderson, S., Parbery-Clark, A., White-Schwoch, T., & Kraus, N. (2012). Aging affects neural precision of speech encoding. *Journal of Neuroscience, 32*(41), 14156–14164. <https://doi.org/10.1523/JNEUROSCI.2176-12.2012>, PubMed: 23055485
- Anderson, S., Parbery-Clark, A., White-Schwoch, T., & Kraus, N. (2015). Development of subcortical speech representation in human infants. *Journal of the Acoustical Society of America, 137*(6), 3346–3355. <https://doi.org/10.1121/1.4921032>, PubMed: 26093424
- Barnet, A. B., & Lodge, A. (1967). Click evoked EEG responses in normal and developmentally retarded infants. *Nature, 214*(5085), 252–255. <https://doi.org/10.1038/214252a0>, PubMed: 4226768
- Barnet, A. B., Ohlrich, E. S., Weiss, I. P., & Shanks, B. (1975). Auditory evoked potentials during sleep in normal children from ten days to three years of age. *Electroencephalography and Clinical Neurophysiology, 39*(1), 29–41. [https://doi.org/10.1016/0013-4694\(75\)90124-8](https://doi.org/10.1016/0013-4694(75)90124-8), PubMed: 50197
- Best, C. T., Goldstein, L. M., Nam, H., & Tyler, M. D. (2016). Articulating what infants attune to in native speech. *Ecological Psychology, 28*(4), 216–261. <https://doi.org/10.1080/10407413.2016.1230372>, PubMed: 28367052
- Best, C. T., McRoberts, G. W., LaFleur, R., & Silver-Isenstadt, J. (1995). Divergent developmental patterns for infants' perception of two nonnative consonant contrasts. *Infant Behavior & Development, 18*(3), 339–350. [https://doi.org/10.1016/0163-6383\(95\)90022-5](https://doi.org/10.1016/0163-6383(95)90022-5)
- Bidelman, G. M. (2018). Subcortical sources dominate the neuroelectric auditory frequency-following response to speech. *NeuroImage, 175*, 56–69. <https://doi.org/10.1016/j.neuroimage.2018.03.060>, PubMed: 29604459
- Bidelman, G. M., Davis, M. K., & Pridgen, M. H. (2018). Brainstem-cortical functional connectivity for speech is differentially challenged by noise and reverberation. *Hearing Research, 367*, 149–160. <https://doi.org/10.1016/j.heares.2018.05.018>, PubMed: 29871826
- Chang, E. F., Rieger, J. W., Johnson, K., Berger, M. S., Barbaro, N. M., & Knight, R. T. (2010). Categorical speech representation in human superior temporal gyrus. *Nature Neuroscience, 13*(11), 1428–1432. <https://doi.org/10.1038/nn.2641>, PubMed: 20890293
- Chao, Y. R. (1930). ə sistim əv “toun-letəz” [A system of tone letters]. *Le Maître Phonétique, 8*, 24–27.
- Cheng, Y. Y., Wu, H. C., Tzeng, Y. L., Yang, M. T., Zhao, L. L., & Lee, C. Y. (2013). The development of mismatch responses to Mandarin lexical tones in early infancy. *Developmental Neuropsychology, 38*(5), 281–300. <https://doi.org/10.1080/87565641.2013.799672>, PubMed: 23862633
- Cheour, M., Cehoniene, R., Lehtokoski, A., Luuk, A., Allik, J., Alho, K., & Näätänen, R. (1998). Development of language-specific phoneme representations in the infant brain. *Nature Neuroscience, 1*(5), 351–353. <https://doi.org/10.1038/1561>, PubMed: 10196522

- Cheour-Luhtanen, M., Alho, K., Kujala, T., Sainio, K., Reinikainen, K., Renlund, M., Aaltonen, O., Eerola, O., & Näätänen, R. (1995). Mismatch negativity indicates vowel discrimination in newborns. *Hearing Research*, *82*(1), 53–58. [https://doi.org/10.1016/0378-5955\(94\)00164-L](https://doi.org/10.1016/0378-5955(94)00164-L), PubMed: 7744713
- Coffey, E. B. J., Nicol, T., White-Schwoch, T., Chandrasekaran, B., Krizman, J., Skoe, E., Zatorre, R. J., & Kraus, N. (2019). Evolving perspectives on the sources of the frequency-following response. *Nature Communications*, *10*, Article 5036. <https://doi.org/10.1038/s41467-019-13003-w>, PubMed: 31695046
- Conboy, B. T., & Kuhl, P. K. (2011). Impact of second-language experience in infancy: Brain measures of first- and second-language speech perception. *Developmental Science*, *14*(2), 242–248. <https://doi.org/10.1111/j.1467-7687.2010.00973.x>, PubMed: 21499512
- Cone, B., & Whitaker, R. (2013). Dynamics of infant cortical auditory evoked potentials (CAEPs) for tone and speech tokens. *International Journal of Pediatric Otorhinolaryngology*, *77*(7), 1162–1173. <https://doi.org/10.1016/j.ijporl.2013.04.030>, PubMed: 23722003
- Cortesa, C. S., Hudac, C. M., & Molfese, D. L. (2019). Dynamic effects of habituation and novelty detection on newborn event-related potentials. *Brain and Language*, *199*, Article 104695. <https://doi.org/10.1016/j.bandl.2019.104695>, PubMed: 31610478
- Delorme, A., & Makeig, S. (2004). EEGLAB: An open source toolbox for analysis of single-trial EEG dynamics including independent component analysis. *Journal of Neuroscience Methods*, *134*(1), 9–21. <https://doi.org/10.1016/j.jneumeth.2003.10.009>, PubMed: 15102499
- Eilers, P. H. C. (2003). A perfect smoother. *Analytical Chemistry*, *75*(14), 3631–3636. <https://doi.org/10.1021/ac034173t>, PubMed: 14570219
- Escera, C. (2017). The role of the auditory brainstem in regularity encoding and deviance detection. In N. Kraus, S. Anderson, T. White-Schwoch, R. R. Fay, & A. N. Popper (Eds.), *The frequency-following response: A window into human communication* (pp. 101–120). Springer. [https://doi.org/10.1007/978-3-319-47944-6\\_5](https://doi.org/10.1007/978-3-319-47944-6_5)
- Ettlinger, M., Bradlow, A. R., & Wong, P. C. M. (2014). Variability in the learning of complex morphophonology. *Applied Psycholinguistics*, *35*(4), 807–831. <https://doi.org/10.1017/S0142716412000586>
- Fromkin, V. (2014). *Tone: A linguistic survey*. Academic Press.
- Gansonre, C., Højlund, A., Leminen, A., Bailey, C., & Shtyrov, Y. (2018). Task-free auditory EEG paradigm for probing multiple levels of speech processing in the brain. *Psychophysiology*, *55*(11), Article e13216. <https://doi.org/10.1111/psyp.13216>, PubMed: 30101984
- Godey, B., Schwartz, D., De Graaf, J. B., Chauvel, P., & Liégeois-Chauvel, C. (2001). Neuromagnetic source localization of auditory evoked fields and intracerebral evoked potentials: A comparison of data in the same patients. *Clinical Neurophysiology*, *112*(10), 1850–1859. [https://doi.org/10.1016/S1388-2457\(01\)00636-8](https://doi.org/10.1016/S1388-2457(01)00636-8), PubMed: 11595143
- Government of Hong Kong SAR. (2016). Hong Kong: The Facts. <https://www.gov.hk/en/about/about/hk/facts.htm>
- Grimaldi, M. (2018). The phonetics-phonology relationship in the neurobiology of language. In R. Petrosino, P. Cerrone, & H. van der Hulst (Eds.), *From sounds to structures* (pp. 65–103). De Gruyter. <https://doi.org/10.1515/9781501506734-003>
- Hartley, D. E. H., & King, A. J. (2010). Development of the auditory pathway. In A. Rees & A. R. Palmer (Eds.), *The Oxford handbook of auditory science: The auditory brain* (pp. 361–386). Oxford University Press.
- Hartley, D. E. H., Vongpaisal, T., Xu, J., Shepherd, R. K., King, A. J., & Isaiah, A. (2010). Bilateral cochlear implantation in the ferret: A novel animal model for behavioral studies. *Journal of Neuroscience Methods*, *190*(2), 214–228. <https://doi.org/10.1016/j.jneumeth.2010.05.014>, PubMed: 20576507
- Ingvalson, E. M., & Wong, P. C. M. (2013). Training to improve language outcomes in cochlear implant recipients. *Frontiers in Psychology*, *4*, Article 263. <https://doi.org/10.3389/fpsyg.2013.00263>, PubMed: 23675364
- Intartaglia, B., White-Schwoch, T., Meunier, C., Roman, S., Kraus, N., & Schön, D. (2016). Native language shapes automatic neural processing of speech. *Neuropsychologia*, *89*, 57–65. <https://doi.org/10.1016/j.neuropsychologia.2016.05.033>, PubMed: 27263123
- Jeng, F.-C., Hu, J., Dickman, B., Montgomery-Reagan, K., Tong, M., Wu, G., & Lin, C.-D. (2011). Cross-linguistic comparison of frequency-following responses to voice pitch in American and Chinese neonates and adults. *Ear and Hearing*, *32*(6), 699–707. <https://doi.org/10.1097/AUD.0b013e31821cc0df>, PubMed: 21543983
- Jeng, F.-C., Lin, C.-D., Chou, M. S., Hollister, G. R., Sabol, J. T., Mayhugh, G. N., Wang, T.-C., & Wang, C.-Y. (2016). Development of subcortical pitch representation in three-month-old Chinese infants. *Perceptual and Motor Skills*, *122*, 123–135. <https://doi.org/10.1177/0031512516631054>, PubMed: 27420311
- Jeng, F.-C., Peris, K. S., Hu, J., & Lin, C.-D. (2013). Evaluation of an automated procedure for detecting frequency-following responses in American and Chinese neonates. *Perceptual and Motor Skills*, *116*, 456–465. <https://doi.org/10.2466/24.10.PMS.116.2.456-465>, PubMed: 24032322
- Key, A. P., Lambert, E. W., Aschner, J. L., & Maitre, N. L. (2012). Influence of gestational age and postnatal age on speech sound processing in NICU infants. *Psychophysiology*, *49*(5), 720–731. <https://doi.org/10.1111/j.1469-8986.2011.01353.x>, PubMed: 22332725
- Krishnan, A., Gandour, J. T., & Bidelman, G. M. (2010). The effects of tone language experience on pitch processing in the brainstem. *Journal of Neurolinguistics*, *23*(1), 81–95. <https://doi.org/10.1016/j.jneuroling.2009.09.001>, PubMed: 20161561
- Krishnan, A., Swaminathan, J., & Gandour, J. T. (2009). Experience-dependent enhancement of linguistic pitch representation in the brainstem is not specific to a speech context. *Journal of Cognitive Neuroscience*, *21*(6), 1092–1105. <https://doi.org/10.1162/jocn.2009.21077>, PubMed: 18702588
- Krishnan, A., Xu, Y., Gandour, J., & Cariani, P. (2005). Encoding of pitch in the human brainstem is sensitive to language experience. *Cognitive Brain Research*, *25*(1), 161–168. <https://doi.org/10.1016/j.cogbrainres.2005.05.004>, PubMed: 15935624
- Kuhl, P. K. (2004). Early language acquisition: Cracking the speech code. *Nature Reviews Neuroscience*, *5*(11), 831–843. <https://doi.org/10.1038/nrn1533>, PubMed: 15496861
- Kuhl, P. K. (2010). Brain mechanisms in early language acquisition. *Neuron*, *67*(5), 713–727. <https://doi.org/10.1016/j.neuron.2010.08.038>, PubMed: 20826304
- Kuhl, P. K. (2014). Early language learning and the social brain. *Cold Spring Harbor Symposia on Quantitative Biology*, *79*, 211–220. <https://doi.org/10.1101/sqb.2014.79.024802>, PubMed: 25943768
- 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). *Philosophical Transactions of the Royal Society B: Biological Sciences*, 363(1493), 979–1000. <https://doi.org/10.1098/rstb.2007.2154>, PubMed: 17846016
- Kushnerenko, E., Čeponienė, R., Balan, P., Fellman, V., Huotilainen, M., & Näätänen, R. (2002). Maturation of the auditory event-related potentials during the first year of life. *NeuroReport: For Rapid Communication of Neuroscience Research*, 13(1), 47–51. <https://doi.org/10.1097/00001756-200201210-00014>, PubMed: 11924892
- Kuuluvainen, S., Nevalainen, P., Sorokin, A., Mittag, M., Partanen, E., Putkinen, V., Seppänen, M., Kähkönen, S., & Kujala, T. (2014). The neural basis of sublexical speech and corresponding nonspeech processing: A combined EEG-MEG study. *Brain and Language*, 130, 19–32. <https://doi.org/10.1016/j.bandl.2014.01.008>, PubMed: 24576806
- Leaver, A. M., & Rauschecker, J. P. (2016). Functional topography of human auditory cortex. *Journal of Neuroscience*, 36(4), 1416–1428. <https://doi.org/10.1523/JNEUROSCI.0226-15.2016>, PubMed: 26818527
- Lippé, S., Martínez-Montes, E., Arcand, C., & Lassonde, M. (2009). Electrophysiological study of auditory development. *Neuroscience*, 164(3), 1108–1118. <https://doi.org/10.1016/j.neuroscience.2009.07.066>, PubMed: 19665050
- Lopez-Calderon, J., & Luck, S. J. (2014). ERPLAB: An open-source toolbox for the analysis of event-related potentials. *Frontiers in Human Neuroscience*, 8, Article 213. <https://doi.org/10.3389/fnhum.2014.00213>, PubMed: 24782741
- Maggu, A. R., Liu, F., Antoniou, M., & Wong, P. C. M. (2016). Neural correlates of indicators of sound change in Cantonese: Evidence from cortical and subcortical processes. *Frontiers in Human Neuroscience*, 10, Article 652. <https://doi.org/10.3389/fnhum.2016.00652>, PubMed: 28066218
- Maggu, A. R., Zong, W., Law, V., & Wong, P. C. M. (2018). Learning two tone languages enhances the brainstem encoding of lexical tones. In *19th annual conference of the international speech communication association, INTERSPEECH 2018* (Hyderabad, India, pp. 1437–1441). Curran Associates. <https://doi.org/10.21437/Interspeech.2018-2130>
- Malmierca, M. S., Carbajal, G. V., & Escera, C. (2019). Deviance detection and encoding acoustic regularity in the auditory mid-brain. In K. Kandler (Ed.), *Oxford handbook of the auditory brainstem* (pp. 706–740). Oxford University Press. <https://doi.org/10.1093/oxfordhob/9780190849061.013.19>
- Mattock, K., & Burnham, D. (2006). Chinese and English infants' tone perception: Evidence for perceptual reorganization. *Infancy*, 10(3), 241–265. [https://doi.org/10.1207/s15327078in1003\\_3](https://doi.org/10.1207/s15327078in1003_3)
- McCarthy, K. M., Skoruppa, K., & Iverson, P. (2019). Development of neural perceptual vowel spaces during the first year of life. *Scientific Reports*, 9, Article 19592. <https://doi.org/10.1038/s41598-019-55085-y>, PubMed: 31862999
- Mesgarani, N., Cheung, C., Johnson, K., & Chang, E. F. (2014). Phonetic feature encoding in human superior temporal gyrus. *Science*, 343(6174), 1006–1010. <https://doi.org/10.1126/science.1245994>, PubMed: 24482117
- Moore, J. K., & Guan, Y. L. (2001). Cytoarchitectural and axonal maturation in human auditory cortex. *Journal of the Association for Research in Otolaryngology*, 2(4), 297–311. <https://doi.org/10.1007/s101620010052>, PubMed: 11833605
- Moore, J. K., Guan, Y. L., & Shi, S. R. (1997). Axogenesis in the human fetal auditory system, demonstrated by neurofilament immunohistochemistry. *Anatomy and Embryology*, 195(1), 15–30. <https://doi.org/10.1007/s004290050021>, PubMed: 9006712
- Moore, J. K., Guan, Y. L., & Shi, S. R. (1998). MAP2 expression in developing dendrites of human brainstem auditory neurons. *Journal of Chemical Neuroanatomy*, 16(1), 1–15. [https://doi.org/10.1016/S0891-0618\(98\)00057-X](https://doi.org/10.1016/S0891-0618(98)00057-X), PubMed: 9924969
- Moore, J. K., & Linthicum, F. H., Jr. (2001). Myelination of the human auditory nerve: Different time courses for Schwann cell and glial myelin. *Annals of Otology, Rhinology, & Laryngology*, 110(7, Pt. 1), 655–661. <https://doi.org/10.1177/000348940111000711>, PubMed: 11465825
- Moore, J. K., & Linthicum, F. H., Jr. (2007). The human auditory system: A timeline of development. *International Journal of Audiology*, 46(9), 460–478. <https://doi.org/10.1080/14992020701383019>, PubMed: 17828663
- Moore, J. K., Perazzo, L. M., & Braun, A. (1995). Time course of axonal myelination in the human brainstem auditory pathway. *Hearing Research*, 87(1–2), 21–31. [https://doi.org/10.1016/0378-5955\(95\)00073-D](https://doi.org/10.1016/0378-5955(95)00073-D), PubMed: 8567438
- Moushegian, G., Rupert, A. L., & Stillman, R. D. (1973). Scalp-recorded early responses in man to frequencies in the speech range. *Electroencephalography & Clinical Neurophysiology*, 35(6), 665–667. [https://doi.org/10.1016/0013-4694\(73\)90223-X](https://doi.org/10.1016/0013-4694(73)90223-X), PubMed: 4128165
- Näätänen, R., Lehtokoski, A., Lennes, M., Cheour, M., Huotilainen, M., Iivonen, A., Vainio, M., Alku, P., Ilmoniemi, R. J., Luuk, A., Allik, J., Sinkkonen, J., & Alho, K. (1997). Language-specific phoneme representations revealed by electric and magnetic brain responses. *Nature*, 385(6615), 432–434. <https://doi.org/10.1038/385432a0>, PubMed: 9009189
- Näätänen, R., & Picton, T. W. (1987). The N1 wave of the human electric and magnetic response to sound: A review and an analysis of the component structure. *Psychophysiology*, 24(4), 375–425. <https://doi.org/10.1111/j.1469-8986.1987.tb00311.x>, PubMed: 3615753
- Näätänen, R., & Winkler, I. (1999). The concept of auditory stimulus representation in cognitive neuroscience. *Psychological Bulletin*, 125(6), 826–859. <https://doi.org/10.1037/0033-2909.125.6.826>, PubMed: 10589304
- Novak, G. P., Kurtzberg, D., Kreuzer, J. A., & Vaughan, H. G. (1989). Cortical responses to speech sounds and their formants in normal infants: Maturation sequence and spatiotemporal analysis. *Electroencephalography & Clinical Neurophysiology*, 73(4), 295–305. [https://doi.org/10.1016/0013-4694\(89\)90108-9](https://doi.org/10.1016/0013-4694(89)90108-9), PubMed: 2477216
- Ortiz-Mantilla, S., Hamalainen, J. A., Musacchia, G., & Benasich, A. A. (2013). Enhancement of gamma oscillations indicates preferential processing of native over foreign phonemic contrasts in infants. *Journal of Neuroscience*, 33(48), 18746–18754. <https://doi.org/10.1523/JNEUROSCI.3260-13.2013>, PubMed: 24285881
- 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 of the United States of America*, 110(37), 15145–15150. <https://doi.org/10.1073/pnas.1302159110>, PubMed: 23980148
- Ponton, C. W., Eggermont, J. J., Khosla, D., Kwong, B., & Don, M. (2002). Maturation of human central auditory system activity: Separating auditory evoked potentials by dipole source modeling. *Clinical Neurophysiology*, 113(3), 407–420. [https://doi.org/10.1016/S1388-2457\(01\)00733-7](https://doi.org/10.1016/S1388-2457(01)00733-7), PubMed: 11897541



- Raznahan, A., Shaw, P. W., Lerch, J. P., Clasen, L. S., Greenstein, D., Berman, R., Pipitone, J., Chakravarty, M. M., & Giedd, J. N. (2014). Longitudinal four-dimensional mapping of subcortical anatomy in human development. *Proceedings of the National Academy of Sciences of the United States of America*, *111*(4), 1592–1597. <https://doi.org/10.1073/pnas.1316911111>, PubMed: 24474784
- Reetzke, R., Xie, Z., Llanos, F., & Chandrasekaran, B. (2018). Tracing the trajectory of sensory plasticity across different stages of speech learning in adulthood. *Current Biology*, *28*(9), 1419–1427. <https://doi.org/10.1016/j.cub.2018.03.026>, PubMed: 29681473
- Ribas-Prats, T., Almeida, L., Costa-Faidella, J., Plana, M., Corral, M. J., Gómez-Roig, M. D., & Escera, C. (2019). The frequency-following response (FFR) to speech stimuli: A normative dataset in healthy newborns. *Hearing Research*, *371*, 28–39. <https://doi.org/10.1016/j.heares.2018.11.001>, PubMed: 30448690
- Rivera-Gaxiola, M., Klarman, C. A. L., Garcia-Sierra, A., & Kuhl, P. K. (2005). Neural patterns to speech and vocabulary growth in American infants. *Neuroreport*, *16*(5), 495–498. <https://doi.org/10.1097/00001756-200504040-00015>, PubMed: 15770158
- Rivera-Gaxiola, M., Silva-Pereyra, J., & Kuhl, P. K. (2005). Brain potentials to native and non-native speech contrasts in 7- and 11-month-old American infants. *Developmental Science*, *8*, 162–172. <https://doi.org/10.1111/j.1467-7687.2005.00403.x>, PubMed: 15720374
- Romand, R., & Romand, M.-R. (1982). Myelination kinetics of spiral ganglion cells in kitten. *Journal of Comparative Neurology*, *204*(1), 1–5. <https://doi.org/10.1002/cne.902040102>, PubMed: 7056884
- Romand, R., Sans, A., Romand, M. R., & Marty, R. (1976). The structural maturation of the stato-acoustic nerve in the cat. *Journal of Comparative Neurology*, *170*(1), 1–15. <https://doi.org/10.1002/cne.901700102>, PubMed: 977813
- Ross, B., & Tremblay, K. (2009). Stimulus experience modifies auditory neuromagnetic responses in young and older listeners. *Hearing Research*, *248*(1–2), 48–59. <https://doi.org/10.1016/j.heares.2008.11.012>, PubMed: 19110047
- Sano, M., Kaga, K., Kuan, C. C., Ino, K., & Mima, K. (2007). Early myelination patterns in the brainstem auditory nuclei and pathway: MRI evaluation study. *International Journal of Pediatric Otorhinolaryngology*, *71*(7), 1105–1115. <https://doi.org/10.1016/j.ijporl.2007.04.002>, PubMed: 17485121
- Shafer, V. L., Yu, Y. H., & Wagner, M. (2015). Maturation of cortical auditory evoked potentials (CAEPs) to speech recorded from frontocentral and temporal sites: Three months to eight years of age. *International Journal of Psychophysiology*, *95*(2), 77–93. <https://doi.org/10.1016/j.ijpsycho.2014.08.1390>, PubMed: 25219893
- Shiga, T., Althen, H., Cornella, M., Zarnowicz, K., Yabe, H., & Escera, C. (2015). Deviance-related responses along the auditory hierarchy: Combined FFR, MLR and MMN evidence. *PLOS One*, *10*(9), Article e0136794. <https://doi.org/10.1371/journal.pone.0136794>, PubMed: 26348628
- Shtyrov, Y., Kujala, T., Palva, S., Ilmoniemi, R. J., & Näätänen, R. (2000). Discrimination of speech and of complex nonspeech sounds of different temporal structure in the left and right cerebral hemispheres. *NeuroImage*, *12*(6), 657–663. <https://doi.org/10.1006/nimg.2000.0646>, PubMed: 11112397
- Skoe, E., & Kraus, N. (2010). Auditory brain stem response to complex sounds: A tutorial. *Ear and Hearing*, *31*(3), 302–324. <https://doi.org/10.1097/AUD.0b013e3181c8db272>, PubMed: 20084007
- Skoe, E., Krizman, J., Anderson, S., & Kraus, N. (2015). Stability and plasticity of auditory brainstem function across the lifespan. *Cerebral Cortex*, *25*(6), 1415–1426. <https://doi.org/10.1093/cercor/bht311>, PubMed: 24366906
- Slabu, L., Grimm, S., & Escera, C. (2012). Novelty detection in the human auditory brainstem. *Journal of Neuroscience*, *32*(4), 1447–1452. <https://doi.org/10.1523/JNEUROSCI.2557-11.2012>, PubMed: 22279229
- Small, S. A., Sharma, M., Bradford, M., & Vasuki, P. R. M. (2018). The effect of signal to noise ratio on cortical auditory-evoked potentials elicited to speech stimuli in infants and adults with normal hearing. *Ear and Hearing*, *39*(2), 305–317. <https://doi.org/10.1097/AUD.0000000000000487>, PubMed: 28863034
- Song, J. H., Skoe, E., Wong, P. C., & Kraus, N. (2008). Plasticity in the adult human auditory brainstem following short-term linguistic training. *Journal of Cognitive Neuroscience*, *20*(10), 1892–1902. <https://doi.org/10.1162/jocn.2008.20131>, PubMed: 18370594
- Suga, N. (2008). Role of corticofugal feedback in hearing. *Journal of Comparative Physiology A*, *194*(2), 169–183. <https://doi.org/10.1007/s00359-007-0274-2>, PubMed: 18228080
- Suga, N., Gao, E., Zhang, Y., Ma, X., & Olsen, J. F. (2000). The corticofugal system for hearing: Recent progress. *Proceedings of the National Academy of Sciences of the United States of America*, *97*(22), 11807–11814. <https://doi.org/10.1073/pnas.97.22.11807>, PubMed: 11050213
- Tallon-Baudry, C., Bertrand, O., Delpuech, C., & Pernier, J. (1996). Stimulus specificity of phase-locked and non-phase-locked 40 Hz visual responses in human. *Journal of Neuroscience*, *16*(13), 4240–4249. <https://doi.org/10.1523/JNEUROSCI.16-13-04240.1996>, PubMed: 8753885
- Tichko, P., & Skoe, E. (2017). Frequency-dependent fine structure in the frequency-following response: The byproduct of multiple generators. *Hearing Research*, *348*, 1–15. <https://doi.org/10.1016/j.heares.2017.01.014>, PubMed: 28137699
- Tremblay, K. L., & Kraus, N. (2002). Auditory training induces asymmetrical changes in cortical neural activity. *Journal of Speech, Language, and Hearing Research*, *45*(3), 564–572. [https://doi.org/10.1044/1092-4388\(2002\)045](https://doi.org/10.1044/1092-4388(2002)045)
- Tremblay, K. L., Kraus, N., & McGee, T. (1998). The time course of auditory perceptual learning: Neurophysiological changes during speech-sound training. *Neuroreport*, *9*(16), 3557–3560. <https://doi.org/10.1097/00001756-199811160-00003>, PubMed: 9858359
- Tremblay, K. L., Kraus, N., McGee, T., Ponton, C. W., & Otis, A. B. (2001). Central auditory plasticity: Changes in the N1-P2 complex after speech-sound training. *Ear and Hearing*, *22*(2), 79–90. <https://doi.org/10.1097/00003446-200104000-00001>, PubMed: 11324846
- Tremblay, K. L., Ross, B., Inoue, K., McClannahan, K., & Collet, G. (2014). Is the auditory evoked P2 response a biomarker of learning? *Frontiers in Systems Neuroscience*, *8*, Article 28. <https://doi.org/10.3389/fnsys.2014.00028>, PubMed: 24600358
- Werker, J. F., & Tees, R. C. (1984). Cross-language speech perception: Evidence for perceptual reorganization during the first year of life. *Infant Behavior & Development*, *7*(1), 49–63. [https://doi.org/10.1016/S0163-6383\(84\)80022-3](https://doi.org/10.1016/S0163-6383(84)80022-3)
- Werker, J. F., Yeung, H. H., & Yoshida, K. A. (2012). How do infants become experts at native-speech perception? *Current Directions in Psychological Science*, *21*, 221–226. <https://doi.org/10.1177/0963721412449459>
- Wierenga, L., Langen, M., Ambrosino, S., van Dijk, S., Oranje, B., & Durston, S. (2014). Typical development of basal ganglia,

- hippocampus, amygdala and cerebellum from age 7 to 24. *NeuroImage*, 96, 67–72. <https://doi.org/10.1016/j.neuroimage.2014.03.072>, PubMed: 24705201
- Wong, P. C. M., Skoe, E., Russo, N. M., Dees, T., & Kraus, N. (2007). Musical experience shapes human brainstem encoding of linguistic pitch patterns. *Nature Neuroscience*, 10(4), 420–422. <https://doi.org/10.1038/nn1872>, PubMed: 17351633
- Wunderlich, J. L., & Cone-Wesson, B. K. (2006). Maturation of CAEP in infants and children: A review. *Hearing Research*, 212(1–2), 212–223. <https://doi.org/10.1016/j.heares.2005.11.008>, PubMed: 16480841
- Yeung, H. H., Chen, K. H., & Werker, J. F. (2013). When does native language input affect phonetic perception? The precocious case of lexical tone. *Journal of Memory and Language*, 68, 123–139. <https://doi.org/10.1016/j.jml.2012.09.004>
- Yip, M. (2002). *Tone*. Cambridge University Press. <https://doi.org/10.1017/CBO9781139164559>
- Zhao, T. C., & Kuhl, P. K. (2018). Linguistic effect on speech perception observed at the brainstem. *Proceedings of the National Academy of Sciences of the United States of America*, 115(35), 8716–8721. <https://doi.org/10.1073/pnas.1800186115>, PubMed: 30104356