Manuscript title: Neural Speech Encoding in Infancy Predicts Future Language and Communication Difficulties

Short title: Infant Neural Features Predict Future Language **Authors:**

Patrick C. M. Wong, PhD, CCC-SLP, Brain and Mind Institute and Department of Linguistics and Modern Languages, The Chinese University of Hong Kong, Hong Kong SAR, China

Ching Man Lai, PhD, Brain and Mind Institute, The Chinese University of Hong Kong, Hong Kong SAR, China,

Peggy H. Y. Chan, PhD, Brain and Mind Institute and Department of Paediatrics, The Chinese University of Hong Kong, Hong Kong SAR, China

Ting Fan Leung, MD, Department of Paediatrics, The Chinese University of Hong Kong, Hong Kong SAR, China

Hugh Simon Lam, MD, Department of Paediatrics, The Chinese University of Hong Kong, Hong Kong SAR, China

Gangyi Feng, PhD, Brain and Mind Institute and Department of Linguistics and Modern Languages, The Chinese University of Hong Kong, Hong Kong SAR, China

Akshay R. Maggu, PhD, Brain and Mind Institute, Department of Linguistics and Modern Languages, The Chinese University of Hong Kong, Hong Kong SAR, China, and Department of Psychology and Neuroscience, Duke University, Durham, NC, USA

Nikolay Novitskiy, PhD, Brain and Mind Institute and Department of Linguistics and Modern Languages, The Chinese University of Hong Kong, Hong Kong SAR, China

Corresponding and reprint request author:
Patrick CM Wong, PhD
Dept. of Linguistics & Modern Languages, The Chinese University of Hong Kong
Room G03, Leung Kau Kui Building, Shatin, N.T., Hong Kong SAR, China
+852 3943 7009 (tel), p.wong@cuhk.edu.hk (e-mail)

Keywords: Language impairment, language development, brain development, Auditory brainstem response, early detection, early intervention

Conflict of interests: P.W., N.N., A.M., H.S.L., G.F., and T.F.L. are co-inventors of a patent application associated with the research reported here. P.W. is the founder of a startup company, Foresight Language and Learning Solutions Limited, supported by a Hong Kong SAR government tech-company startup scheme for universities.

Sources of financial assistance: This study was supported by the Innovation and Technology Fund of the Hong Kong SAR Government (ITS/067/18), including funding from its research talent program (InP/285/19, InP/286/19, PiH/030/19, and PiH/034/19). Support from the Dr. Stanley Ho Medical Development Foundation for the Stanley Ho Developmental Cohort Study is also acknowledged.

Abstract

Purpose: To construct an objective, and cost-effective prognostic tool to forecast the future language and communication abilities of individual infants.

Method: Speech-evoked electroencephalography (EEG) data were collected from 118 infants during the first year of life during the exposure to speech stimuli that differed principally in fundamental frequency. Language and communication outcomes, namely four subtests of the MacArthur-Bates Communicative Development Inventories (MCDI)—Chinese version, were collected between 3 to 16 months after initial EEG testing. In the two-way classification, children were classified into those with future MCDI scores below the 25th percentile for their age group and those above the same percentile, while the three-way classification classified them into < 25th, 25th – 75th, and > 75th percentile groups. Machine learning (support vector machine classification) with cross-validation was used for model construction. Statistical significance was assessed.

Results: Across the four MCDI measures of early gestures, later gestures, vocabulary comprehension, and vocabulary production, the areas under the receiver-operating characteristic curve (AUC) of the predictive models were respectively .92 ± .031, .91 ± .028, .90 ± .035, and .89 ± .039 for the two-way classification, and .88 ± .041, .89 ± .033, .85 ± .047 and .85 ± .050 for the three-way classification (p <.01 for all models).

Conclusions: Future language and communication variability can be predicted by an objective EEG method that indicates the function of the auditory neural pathway foundational to spoken

language development, with precision sufficient for individual predictions. Longer-term research is needed to assess predictability of categorical diagnostic status.

List of abbreviations: area under the receiver-operating characteristic curve (AUC), electroencephalography (EEG), early intervention (EI), frequency following response (FFR), MacArthur-Bates Communicative Development Inventories (MCDI), long-latency response (LLR), socioeconomic status (SES)

Introduction

Poor language skills can place a long-term burden on both individuals and the society, owing to its links to mental health and behavioral adjustment (Bornstein, 2013), academic achievement, and employment (Beiser & Hou, 2001; Bleakley & Chin, 2004). Estimates of prevalence of language delay and impairment vary by studies and age (Law, Boyle, Harris, Harkness, & Nye, 2000), but they can be as high as asthma (about 7%, Centers for Disease Control and Prevention, 2019) and even childhood obesity (about 17%, Sanyaolu, Okorie, Qi, Locke, & Rehman, 2019). These estimates are much higher than the prevalence data for Autism Spectrum Disorder (ASD) (1 out of 54 children in the US) (CDC, 2020). Language and communication features are common deficit domains across most neurodevelopmental disorders (American Psychiatric Association: Diagnostic and Statistical Manual of Mental Disorders, Fifth Edition., 2013). Language impairment, including developmental language disorder, therefore constitutes a public health concern (Law, Reilly, & Snow, 2013). Without intervention, language development remains remarkably stable: children who are ranked low in language and communication ability among their peers at 2 years of age continue to rank low at age 15 years (Bornstein, Hahn, Putnick, &

Pearson, 2018). Part of this developmental stability comes from the child's environment, including interaction between parents and children (Bornstein, Tamis-LeMonda, Hahn, & Haynes, 2008; Bornstein, Tamis-LeMonda, & Haynes, 1999; Sy, Gottfried, & Gottfried, 2013). With early intervention (EI), language can be improved, as the nervous system is the most plastic in the early years (Roberts & Kaiser, 2015; Ingvalson & Wong, 2013). Interventions that are delivered early have higher economic returns than those administered later in life (e.g., through adult literacy programs, Heckman, 2008).

Because language development is stable and because EI is effective, we must develop a method to predict future language abilities at the earliest possible time point. Early prediction would allow caregivers and clinicians to plan for EI to potentially prevent or reduce the severity of language impairment. To date, no method precise enough to predict language functions at the individual child level has been developed. Although large-scale cohort studies have repeatedly found that early language predicts future language (Bornstein, Hahn, & Haynes, 2004; Bornstein, Hahn, & Putnick, 2016b; Bornstein et al., 2018; Bornstein, Hahn, Putnick, & Suwalsky, 2014), these studies were based on group-level results, and their goal was not to construct individual-level predictive models. Behavioral assessment in young children is especially prone to measurement error and dependent on the skill and subjectivity of the assessor (Andersson, 2004; DiPietro & Larson, 1989). It is not surprising that smaller-scale behavioral studies—without the benefit of a larger sample size to correct for measurement errors—have often found less stability in language development (Paul, Looney, & Dahm, 1991; Rescorla, Mirak, & Singh, 2000; Stothard, Snowling, Bishop, Chipchase, & Kaplan, 1998). Even in larger-scale studies where language developmental stability was observed, behavioral assessment and demographic factors only explained a small amount of variance in language outcomes (Reilly et al., 2007).

In this study, we investigate a novel approach to constructing prognostic models for language developmental outcome that are precise enough for making individual-level predictions. In recent years, human neuroscience research has repeatedly demonstrated that neural measures have become sufficiently precise that they can predict cognitive outcomes better than behavioral measures alone, when machine-learning analytics are used (see Gabrieli, Ghosh, & WhitfieldGabrieli, 2015 for a review). The prediction of future language outcomes of hearing-impaired children with cochlear implants, for example, has found specificity of up to 88% (Feng et al., 2018). The native language of the vast majority of children is spoken. Spoken language development depends on the functions of the auditory neural system (e.g., Feng et al., 2018; Deng et al., 2016,

2018), as the process of learning depends on an ability to distinguish fine acoustic details (e.g., Ingvalson et al., 2014). Auditory encoding, in particular, enables incoming speech signals to be accurately represented. It can be measured by neural phase-locking to the frequencies of the speech stimulus via the frequency following response (FFR) in electroencephalography (EEG) testing (e.g., Kraus & Nicol, 2005; Wong, Skoe, Russo, Dees, & Kraus, 2007). FFR measures are associated with a range of language processes, including speech (Thompson et al., 2019) and literacy (White-Schwoch et al., 2015). Measurement of the neural encoding of speech is a prime candidate for predicting future language development in young children. Previous research has provided preliminary evidence for the connection between neural responses to speech (including long-latency responses [LLRs]) during infancy and future language and communication development (Guttorm, Leppänen, Hämäläinen, Eklund, & Lyytinen, 2010; Kuhl et al., 2008; Molfese, 2000). However, attempts to measure this relationship have yet to reach the level of precision for prediction at the individual level. In the present study, we enrolled 118 healthy infants from Cantonese Chinese-speaking families. All underwent an EEG procedure that can be administered while the infant is naturally sleeping. Language and communication outcome was measured using the Words and Gestures form of the MacArthur- Bates Communicative Development Inventories (MCDI)—Chinese version (Tardif & Fletcher, 2008) up to 16 months after initial EEG testing. For measurement of early language functions, the MCDI is among of the most commonly used instruments (Law & Roy, 2008) and its use has also been validated for various clinical populations (Smith & Mirenda, 2009; Thal, Desjardin, & Eisenberg, 2007). Machine-learning techniques with the application of cross-validation were used to construct the predictive models.

Methods

Participants. A community sample of 118 Cantonese-learning infants (57 females) participated in this study (see Table S1 for a summary of demographic information). The participants were enrolled by their parents in our study in response to a local news report about our research and from local hospitals. Despite our ability to reach out to the community, our sample is not entirely a random sample (e.g., we made no attempt to enroll families from different districts). At the time of EEG testing, infants were below 12 months of age (mean = 3.8, range .8 - 12.4, SD = 2.29). Their language and communication outcome was measured between 3 to 16 months after EEG testing, when they were between 8 and 18 months of age (mean = 12.5, range 8 - 18, SD = 1.98) (see Fig. S1 for the age of the participants at EEG and outcome). Parents were native Cantonese speakers and reported speaking Cantonese at home. Although Hong Kong is a multilingual society, the vast majority of the population speaks Cantonese (88.9%) as their dominant language. Only a small minority speak English (4.3%) or Mandarin (1.9%) as the dominant language (Government of Hong Kong SAR, 2016). All participants passed universal hearing screening shortly after birth, and their parents reported no significant developmental or neurological disorders. One hundred and thirteen subjects were born at least 37 weeks after gestation. The remaining 5 subjects were born between 34 and 36.7 weeks after gestation (total statistics for gestational age: mean 38.8 range 34 - 41 SD 1.29 weeks), with a birth weight of at least 2 kg (mean = 3.1 kg, range 2.2 - 4.3 kg, SD =.4 kg). Our convenience sampling approach is noteworthy because it allows us to construct predictive models that are versatile in terms of when EEG and outcome data are collected during the first two years of life. Our model construction techniques take into consideration variability in age (see below). This study was approved by the Joint Chinese University of Hong Kong – New Territories East Cluster Clinical Research Ethics Committee. Informed consent was obtained from the caregivers.

Sample Size Determination. We used a correlational approach to estimate the number of participants needed. As far as we are aware, there is no standard way of calculating power for the kind of study we conducted that used machine learning to link EEG with developmental outcome. The correlation between an EEG measure (Pitch Strength) and outcome (MCDI Later Gestures) was $r = .3554$ for the first 20 participants enrolled. With this effect size, 59 participants would be needed to achieve 80% power given alpha of .05. Because this correlational approach is an imperfect method of estimating sample size and does not consider cross-validation procedures, we believe doubling the minimum requirement (to a final sample of 118) would be sufficiently conservative.

EEG Testing and Outcome Measures. The infants underwent EEG testing while being held by their caregivers (see Fig. S2 for experimental setup). During the test, the infants listened to three speech stimuli: the Native /ga2/, Native /ga4/ and Non-Native /ga3/ (‘2’ and ‘4’ signal the

Cantonese rising and falling tones and ‘3’ signals the Mandarin dipping tone which is lacking in Cantonese). Each syllable was presented 1000 times. Continuous EEG was collected from Ag/AgCl electrodes at Cz, M1 (left mastoid) and M2 (right mastoid) at a 20 kHz sampling rate (Compumedics, Australia) with CPz as a reference and Fpz as a ground. Cz data were re-referenced offline with the average of two mastoids for the subsequent analysis. The Words and Gestures form of the MacArthur-Bates Communicative Development Inventories (MCDI)—Cantonese version (Tardif & Fletcher, 2008) was used to measure language and communication outcome. For infants within our age range, four components were used as outcome measures: early gestures, later gestures, vocabulary comprehension and vocabulary production. Normative scores for each measure based on the child’s age and sex were used. Children were classified into three groups based on their normative scores for each component: < 25th percentile (below average), 25th – 75th percentile (average), and > 75th percentile (above average) in three-way classification. In two-way classification, they were divided into < 25th percentile (below average) and the rest (> 25th percentile).

Analysis. Extraction of EEG measures. The data were processed in two parallel pipelines. The short-latency response (also known as FFR) pipeline included 80 - 1500 Hz bandpass filtering and -50 ms to 250 ms peristimulus epoch extraction. The long-latency response (LLR) pipeline went through .1 - 40 Hz filtering and -100 - 500 ms epoching. This frequency range has been used in other LLR studies (Bishop, Hardiman, Uwer, & Von Suchodoletz, 2007; Paquette et al., 2015; Polich, Aung, & Dalessio, 1988). Both time- and frequency domain measures were extracted separately for each of the three tones. Ultimately, 69 neural measures were derived from the EEG signal (see SI for details).

Model Construction. Fig. S3 shows the model construction procedures including the use of support vector machine classification (Boser, Guyon, & Vapnik, 1992). Models with only non-neural measures were compared with models with both neural and non-neural measures. The details of data analysis and model construction are provided in SI. It is important to highlight that a crossvalidation method is used here, which increases our ability to make predictions about unseen data (see Gabrieli et al., 2015 for a review of its usage in cognitive research).

Blinding. The researcher (N.N.) who analyzed the EEG data and conducted model construction was not involved in directly testing the participants. The researchers who conducted EEG testing (C.M.L. and P.C.) did not analyze the EEG data and would not have known the EEG results, even though they were involved in analyzing the MCDI outcome data.

Results

Short- and Long-Latency Responses to Speech. As a group, the infant participants encoded both native and non-native speech sounds (lexical tones) with a high level of fidelity, indicated by a significant r^2 values (p <.001 for all 3 tones) between the stimulus and response F0 (henceforth "pitch") contours derived from the short-latency responses for Native /ga4/ (.83) and Non-Native /ga3/ (.97) and Native /ga2/ (.31) (see Fig. 1A and 1B for the participants' response waveforms and autocorrelograms respectively. Note how the response generally follows the stimulus frequency plotted as a white line).

------------------------ insert Figure 1 around here -----------------------------------

Pitch Strength is one of the 69 EEG measures, and indicates how well an infant encodes pitch information of the stimulus. An Age x Tone repeated-measures ANOVA of the Pitch Strength revealed the main effects of age (F(1,116) = 7.85, p =.00594, partial η^2 = .063) and Tone (F(2,232) = 14.07, p =.00000171, partial η^2 = .108) but no significant Age x Tone interaction (F(2,232) =.589, p = .556, partial η^2 = .005) (Fig. 1C). The main effect of Tone was not driven by native language status but by larger Pitch Strength of Native /ga2/ encoding (Native /ga2/ vs Non-Native /ga3/ p = .000052, Non-Native /ga3/ vs Native /ga4/ p =.86, Native /ga2/ vs Native /ga4/ p = .000052). Note that the Pitch Strength is quite variable across individual infants, despite their high level of performance as a group. In terms of LLR (Fig. 2A), infants were again quite variable in their responses, but note that LLR signal-to-noise ratio (SNR), also one of the 69 EEG measures, did not grow with age and was not different among the tones. That was revealed by a lack of main effects of Age ($F(1,116) = 1.29$, $p = .258$, partial $\eta^2 = .011$) or Tone ($F(2,232) = 1.50$, $p = .224$, partial $\eta^2 = .013$). Likewise, no Age x Tone interaction was found ($F(2,232) = .70$, $p = .497$, partial $\eta^2 = .006$) (Fig. 2B).

----------------------- insert Figure 2 around here ----------------------------------

Performance of Predictive Models. We report the results of four sets of predictive models here. Each set consists of predicting four MCDI measures (early gestures, later gestures, vocabulary comprehension, and vocabulary production). The four sets of models are three-way classification based on 1) non-neural measures only (sex, birth weight, gestational age), 2) non-neural measures combined with 69 EEG neural measures, in addition to age gap between EEG and outcome assessment, and two-way classification based on 3) non-neural measures only, and 4) non-neural measures combined with 69 EEG neural measures, in addition to age gap between EEG and outcome assessment. As shown in Fig. 1, some of the neural measures grow rapidly with age. Thus, in order to remove the effect of neural maturation, a regression model of each of the neural features against age at EEG testing was built and the neural predictor was the residual of that model. Table 1 summarizes the results. The best models were two-way classification based on both neural and non-neural measures, with the area under the receiver-operating characteristic curve (AUC) of at least .89 for all four MCDI measures. In both two- and three-way classification, models with neural and non-neural measures significantly outperformed models with non-neural measures only (Fig. 3). In addition, we constructed models with participants' family socioeconomic status (SES) as an additional non-neural measure, and these models did not differ from the models without SES (Fig.

S4).

------------------------ insert Figure 3 around here ------------------------------------

Age and Model Performance. A unique feature of our study is that EEG and outcome assessments were conducted over a wide age range, with EEG collected any time before 12 months of age, and language and communication outcomes were taken between 3 and 16 months after EEG testing. An important question is whether model performance is affected by age (e.g., whether EEG taken at an earlier age would have produced better results, and whether prediction of outcome closer to EEG testing would have been more accurate). We found no relationship between either age at EEG acquisition (Fig. 4A) or age gap between EEG and outcome assessment (Fig. 4B) for 3-way classification and a median split of the age variables. This suggests that EEG can be taken at any point during infancy up to 12 months to predict outcome up to 16 months, and that all predictions would be similarly robust (with AUC above 89%).

------------------------ insert Figure 4 around here ------------------------------------

Discussion

During the first year of life, typically developing infants develop their native language rapidly (Kuhl & Rivera-Gaxiola, 2008). Language and communication in the first year of life lays an important foundation for subsequent learning. Without special intervention, language development is quite stable. As a group, children whose language ranked low among their peers during the early years continue to rank low in adolescence (Bornstein et al., 2018). This pattern of language developmental stability has been confirmed in numerous large-scale developmental cohort studies conducted on thousands of children across cultures (Bornstein et al., 2004; Bornstein, Hahn, & Putnick, 2016a; Bornstein et al., 2018; Camaioni & Longobardi, 1995; Longobardi, Spataro, Putnick, & Bornstein, 2016). These studies indicate that language developmental outcome is predictable, albeit only at the group level. To be relevant to parents, tools must be developed to provide prognostic indications at the individual level at the earliest possible time point, so that early intervention can be prescribed in order to prevent or to reduce the severity of a potential language problem.

No such instrument to predict future language development is currently available for infants, as far as we are aware. Although He et al. (2020) used neural data to predict language development, the target population was very preterm infants and the predictive performance for forecasting language was only an AUC of .66. Behavioral assessment may not be reliable enough for making individual-level prediction, especially when it comes to assessment of infants (Andersson, 2004; DiPietro & Larson, 1989; Johnston, Propper, Pudney, & Shields, 2014). In this study, we have developed a method that captures an infant's neural responses to speech and these responses to predict future language via machine learning, including mostly full-term infants. The EEG method has been used in numerous studies (Lau et al., 2020; White-Schwoch et al., 2015; Wong et al., 2007), suggesting that it is safe and reliable. In fact, auditory-evoked EEG is already being used for hearing screening and diagnosis in the form of an Auditory Brainstem Response (ABR) assessment (Eggermont, 2017). Our tool is simply extends its known potential in order to predict language development at the individual-infant level by using a predictive algorithm that we have constructed. Compared to predictive models relying only on non-neural measures such as sex, birth weight and gestational age, the models with the addition of neural measures were much more precise, achieving a performance that places them within the range of many medical screening tools (see Maxim, Niebo, & Utell, 2014 for a review). As SES is a correlate of language development at the group level in English-speaking children (e.g., Fernald, Marchman, & Weisleder, 2013), we assessed SES' potential contribution to prediction in a posthoc manner but did not find its addition to improve the precision of predictive models (Fig. S4).

An important feature of our study is that neural responses to speech sound were collected from children up to 12 months of age. We then used a regression model to statistically adjust the neural input to our predictive models, based on the age at which the neural responses were collected. This procedure enabled us to predict flexibly an infant's future language development with data collected at any time during infancy. The significance for caregivers is that their child could be tested at any convenient time during the first year of life without compromising on the predictive performance of the models. The fact that age at EEG did not meaningfully affect the performance of our predictive models (Fig. 4A) endorses the validity of this approach. As language development is stable (Bornstein et al., 2018), it is not surprising that the predictive performance of our models did not change meaningfully as a function of the time between EEG and outcome assessment (Fig. 4B). Future research is needed to assess whether and how much prediction performance changes for longer-term outcomes. Future research should also examine a larger pool of infants tested at fixed intervals during the first year of life (e.g., 3, 6, or 9 months) in order to have an experimental design that does not depend on a regression model to adjust for age when EEG is collected.

It is not possible for one study to examine all relevant populations. The current study was conducted with Chinese-learning infants with Chinese speech stimuli. It is uncertain whether the same set of stimuli can be used for predicting the language development of children from language environments other than Chinese. With our study as a proof-of-concept starting point, future validation studies with different populations and different speech stimuli should be conducted to assess the broader application of our approach to other populations. Importantly, we would like to note that the speech acoustic feature we tested in this study was lexically contrastive pitch (lexical tone). Whether or not our results apply to non-tone-learning infants requires further testing. Furthermore, whether the same results would be obtained if non-speech stimuli that vary in the same pitch patterns are used should also be investigated in future studies. As a first study, we did not use a feature selection procedure in our machine learning process to assess which FFR and LLR measures were most predictive of language outcome, which future research should

investigate.

Prognostic tools such as one developed here are most clinically relevant if they could guide intervention decisions. In the future, research studies should investigate whether EI guided by the tool developed here would result in better language developmental outcomes.

For caregivers who would like to obtain information about the likely language developmental outcome levels of their infants, they could choose to allow their child to receive the speech-evoked EEG test reported here. In case the predicted outcome level is in the below-average range, they could consider engaging in activities that would optimize child language outcomes, such as learning to implement at home child-directed communicative strategies that have shown to be effective (Roberts & Kaiser, 2015). In the future, the test reported here could be incorporated into a standard ABR hearing screening test (Eggermont, 2017) so that information about hearing sensitivity and probable language outcome could be obtained simultaneously, without the need of administering two separate EEG sessions. The need and public health impact of such a universal screening approach at the population level should be carefully examined (e.g., a cost-benefit analysis may be required).

Conclusions

Language impairment is costly to the individual family affected and to society as a whole. In Australia, for example, the cost of language impairment to society was estimated to be AUS$3.3B per year (Cronin, 2017), which can be as high as the economic burden (health system and productivity losses) of asthma (Deloitte Access Economics, 2015). Waiting for the child to be older to make a reliable, formal diagnosis (Stothard et al., 1998) could mean the loss of years of EI opportunity. Here, we developed a prognostic tool based on EEG methods that are already available (Eggermont, 2017) to predict future language and communication abilities of infants. EEG data can be collected at any time during infancy. The cross-validated models have high specificity and AUC. Our tool developed here is not one that makes a prediction of a categorical clinical diagnosis but rather a prediction of an infant's language abilities on a continuum of functions. A longer-term longitudinal study is needed to ascertain whether a categorical clinical diagnosis can be assigned. With or without a categorical diagnostic label, parents of children who are at any range of ability may still seek the option to obtain predictive information in order to take the appropriate actions to optimize the developmental outcomes for their children.

Acknowledgments

We thank participants of the CUHK Stanley Ho Developmental Cohort Study for participating in this study as well as the Dr. Stanley Ho Medical Development Foundation for their funding support. This study was supported by the Innovation and Technology Fund of the Hong Kong SAR Government (ITS/067/18), including funding from its research talent program (InP/285/19, InP/286/19, PiH/030/19, and PiH/034/19). We also wish to thank Lydia Leung, Kay Wong, Agnes Chou, Mavis Chan, Debby Ma and Charlene Chiu for their assistance with this research, and the CUHK-Utrecht University Joint Centre for Language, Mind and Brain, the CUHK-NTU-WSU Joint Laboratory for Infant Research and the Department of Linguistics and Modern Languages for their support.

References

American Psychiatric Association: Diagnostic and Statistical Manual of Mental Disorders, Fifth Edition. (2013). Arlington, VA: American Psychiatric Association.

Andersson, L. L. (2004). Appropriate and inappropriate interpretation and use of test scores in early intervention. *Journal of Early Intervention*, *27*(1), 55–68. https://doi.org/10.1177/105381510402700105

Beiser, M., & Hou, F. (2001). Language acquisition, unemployment and depressive disorder among Southeast Asian refugees: A 10-year study. *Social Science and Medicine*, *53*(10), 1321–1334. https://doi.org/10.1016/S0277-9536(00)00412-3

Bishop, D. V. M., Hardiman, M., Uwer, R., & Von Suchodoletz, W. (2007). Maturation of the long-latency auditory ERP: Step function changes at start and end of adolescence.

Developmental Science, *10*(5), 565–575. https://doi.org/10.1111/j.1467-7687.2007.00619.x

Bleakley, H., & Chin, A. (2004). Language Skills and Earnings: Evidence from Childhood Immigrants. *Review of Economics and Statistics*, *86*(2), 481–496. https://doi.org/10.1162/003465304323031067

Bornstein, M. H. (2013). Parenting and child mental health: a cross-cultural perspective. *World Psychiatry*, *12*(3), 258–265. https://doi.org/10.1002/wps.20071

Bornstein, M. H., Hahn, C.-S., & Haynes, O. M. (2004). Specific and general language performance across early childhood: Stability and gender considerations. *First Language*, *24*(3), 267–304. https://doi.org/10.1177/0142723704045681

Bornstein, M. H., Hahn, C.-S., & Putnick, D. L. (2016a). Long-term stability of core language skill in children with contrasting language skills. *Developmental Psychology*, *52*(5), 704–716. https://doi.org/10.1037/dev0000111

Bornstein, M. H., Hahn, C.-S., & Putnick, D. L. (2016b). Stability of core language skill across the first decade of life in children at biological and social risk. *Journal of Child Psychology and Psychiatry*, *57*(12), 1434–1443. https://doi.org/10.1111/jcpp.12632

Bornstein, M. H., Hahn, C.-S., Putnick, D. L., & Pearson, R. M. (2018). Stability of core language skill from infancy to adolescence in typical and atypical development. *Science Advances*, *4*(11), eaat7422. https://doi.org/10.1126/sciadv.aat7422

Bornstein, M. H., Hahn, C.-S., Putnick, D. L., & Suwalsky, J. T. D. (2014). Stability of Core Language Skill from Early Childhood to Adolescence: A Latent Variable Approach. *Child Development*, *85*(4), 1346–1356. https://doi.org/10.1111/cdev.12192

Bornstein, M. H., Tamis-LeMonda, C. S., Hahn, C.-S. S., & Haynes, O. M. (2008). Maternal responsiveness to young children at three ages: Longitudinal analysis of a multidimensional, modular, and specific parenting construct. *Developmental Psychology*, *44*(3), 867–874. https://doi.org/10.1037/0012-1649.44.3.867

Bornstein, M. H., Tamis-LeMonda, C. S., & Haynes, O. M. (1999). First words in the second year: Continuity, stability, and models of concurrent and predictive correspondence in vocabulary and verbal responsiveness across age and context. *Infant Behavior and Development*, *22*(1), 65–85. https://doi.org/10.1016/S0163-6383(99)80006-X

Boser, B. E., Guyon, I. M., & Vapnik, V. N. (1992). A training algorithm for optimal margin classifiers. In *Proceedings of the fifth annual workshop on Computational learning theory - COLT '92* (pp. 144–152). New York, New York, USA: ACM Press. https://doi.org/10.1145/130385.130401

Camaioni, L., & Longobardi, E. (1995). Nature and stability of individual differences in early lexical development of Italian-speaking children. *First Language*, *15*(44), 203–218. https://doi.org/10.1177/014272379501504405

Centers for Disease Control and Prevention. (2019). Most Recent Asthma Data. Retrieved from https://www.cdc.gov/asthma/most_recent_national_asthma_data.htm

Centers for Disease Control and Prevention. (2020). Data & Statistics on Autism Spectrum Disorder. Retrieved from https://www.cdc.gov/ncbddd/autism/data.html

Cronin, P. (2017). *The economic impact of Childhood developmental language disorder*. University of Technology Sydney. Retrieved from http://hdl.handle.net/10453/123261

Deloitte Access Economics. (2015). *The Hidden Cost of Asthma*. Kingston ACT, Australia. Retrieved from https://www2.deloitte.com/au/en/pages/economics/articles/hidden-costasthma.html#

Deng, Z., Chandrasekaran, B., Wang, S., & Wong, P. C. M. (2016). Resting-state Low-frequency Fluctuations Reflect Individual Differences in Spoken Language Learning. *Cortex,* 76, 63-78. DOI: http://dx.doi.org/10.1016/j.cortex.2015.11.020.

Deng, Z., Chandrasekaran, B., Wang, S., & Wong, P. C. M. (2018). Training-induced Brain Activation and Functional Connectivity Differentiate Multi-talker and Single-talker Speech Training. *Neurobiology of Learning and Memory*, 151, 1-9. DOI: https://doi.org/10.1016/j.nlm.2018.03.009

DiPietro, J. A., & Larson, S. K. (1989). Examiner effects in the administration of the NBAS: The illusion of reliability. *Infant Behavior and Development*, *12*(1), 119–123. https://doi.org/10.1016/0163-6383(89)90057-X

Eggermont, J. J. (2017). *Hearing loss: Causes, prevention, and treatment*. *Hearing Loss: Causes, Prevention, and Treatment*. London, UK: Academic Press Inc.

Feng, G., Gan Z., Wang, S., Wong, P. C. M., & Chandrasekaran B. (2018). Task-General and Acoustic-invariant Neural Representation of Speech Categories in the Human Brain. *Cerebral Cortex,* 1-14. DOI: https://doi.org/10.1093/cercor/bhx195.

Feng, G., Ingvalson, E. M., Grieco-Calub, T. M., Roberts, M. Y., Ryan, M. E., Birmingham, P., … Wong, P. C. M. (2018). Neural preservation underlies speech improvement from auditory deprivation in young cochlear implant recipients. *Proceedings of the National Academy of Sciences of the United States of America*, *115*(5), E1022–E1031. https://doi.org/10.1073/pnas.1717603115

Fernald, A., Marchman, V. A., & Weisleder, A. (2013). SES differences in language processing skill and vocabulary are evident at 18 months. *Developmental Science*, *16*(2), 234–248. https://doi.org/10.1111/desc.12019

Gabrieli, J. D. E. E., Ghosh, S. S., & Whitfield-Gabrieli, S. (2015). Prediction as a humanitarian and pragmatic contribution from human cognitive neuroscience. *Neuron*, *85*(1), 11–26. https://doi.org/10.1016/j.neuron.2014.10.047

Government of Hong Kong SAR. (2016). Hong Kong: The Facts. Retrieved January 13, 2021, from https://www.gov.hk/en/about/abouthk/facts.htm

Guttorm, T. K., Leppänen, P. H. T., Hämäläinen, J. A., Eklund, K. M., & Lyytinen, H. J. (2010). Newborn Event-Related Potentials Predict Poorer Pre-Reading Skills in Children at Risk for Dyslexia. *Journal of Learning Disabilities*, *43*(5), 391–401. https://doi.org/10.1177/0022219409345005

He, L., Li, H., Wang, J., Chen, M., Gozdas, E., Dillman, J. R., & Parikh, N. A. (2020). A multitask, multi-stage deep transfer learning model for early prediction of neurodevelopment in very preterm infants. *Scientific Reports*, 10, 15072. https://doi.org/10.1038/s41598-020-71914-x

Heckman, J. J. (2008). Schools, skills, and synapses. *Economic Inquiry*, *46*(3), 289–324. https://doi.org/10.1111/j.1465-7295.2008.00163.x

Ingvalson, E. M. & Wong, P. C. M. (2013). Training to Improve Language Outcomes in Cochlear Implant Recipients. *Frontiers in Psychology*, 4(263). DOI: http://doi.org/10.3389/fpsyg.2013.00263.

Ingvalson, E. M., Young, N., Wong, P. C. M. (2014). Auditory-Cognitive Training Improves Language Performance in Prelingually Deafened Cochlear Implant Recipients. International *Journal of Pediatric Otorhinolaryngology*, 78(10), 1624-31. DOI: http://dx.doi.org/10.1016/j.ijporl.2014.07.009.

Johnston, D., Propper, C., Pudney, S., & Shields, M. (2014). Child mental health and educational attainment: multiple observers and the measurement error problem. *Journal Of Applied Econometrics*, *29*, 880–900. https://doi.org/0.1002/jae.2359

Kraus, N., & Nicol, T. (2005). Brainstem origins for cortical 'what' and 'where' pathways in the auditory system. *Trends in Neurosciences*, *28*(4), 176–181. https://doi.org/10.1016/j.tins.2005.02.003

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

Kuhl, P. K., & Rivera-Gaxiola, M. (2008). Neural Substrates of Language Acquisition. *Annual Review of Neuroscience*, *31*(1), 511–534. https://doi.org/10.1146/annurev.neuro.30.051606.094321

Lau, J. C. Y., To, C. K. S., Kwan, J. S. K., Kang, X., Losh, M., & Wong, P. C. M. (2020). Lifelong Tone Language Experience does not Eliminate Deficits in Neural Encoding of Pitch in Autism Spectrum Disorder. *Journal of Autism and Developmental Disorders*, *in press*. https://doi.org/10.1007/s10803-020-04796-7

Law, J., Boyle, J., Harris, F., Harkness, A., & Nye, C. (2000). Prevalence and natural history of primary speech and language delay: findings from a systematic review of the literature.

International Journal of Language & Communication Disorders, *35*(2), 165–188. https://doi.org/10.1080/136828200247133

Law, J., Reilly, S., & Snow, P. C. (2013). Child speech, language and communication need reexamined in a public health context: a new direction for the speech and language therapy profession. *International Journal of Language and Communication Disorders*, *48*(5), 486–496. https://doi.org/10.1111/1460-6984.12027

Law, J., & Roy, P. (2008). Parental report of infant language skills: A review of the development and application of the communicative development inventories. *Child and Adolescent Mental Health*. https://doi.org/10.1111/j.1475-3588.2008.00503.x

Longobardi, E., Spataro, P., Putnick, D. L., & Bornstein, M. H. (2016). Noun and Verb Production in Maternal and Child Language: Continuity, Stability, and Prediction Across the Second Year of Life. *Language Learning and Development*, *12*(2), 183–198. https://doi.org/10.1080/15475441.2015.1048339

Maxim, L. D., Niebo, R., & Utell, M. J. (2014). Screening tests: A review with examples. *Inhalation Toxicology*, *26*(13), 811–828. https://doi.org/10.3109/08958378.2014.955932

Molfese, D. L. (2000). Predicting Dyslexia at 8 Years of Age Using Neonatal Brain Responses. *Brain and Language*, *72*(3), 238–245. https://doi.org/10.1006/brln.2000.2287

Paquette, N., Vannasing, P., Tremblay, J., Lefebvre, F., Roy, M. S., McKerral, M., … Gallagher, A. (2015). Early electrophysiological markers of atypical language processing in prematurely born infants. *Neuropsychologia*, *79*, 21–32. https://doi.org/10.1016/j.neuropsychologia.2015.10.021

Paul, R., Looney, S. S., & Dahm, P. S. (1991). Communication and Socialization Skills at Ages

2 and 3 in “Late-Talking” Young Children. *Journal of Speech, Language, and Hearing Research*, *34*(4), 858–865. https://doi.org/10.1044/jshr.3404.858

Polich, J., Aung, M., & Dalessio, D. J. (1988). Long Latency Auditory Evoked Potentials: Intensity, Inter-Stimulus Interval, and Habituation. *Electroencephalography and Clinical Neurophysiology*, *79*(5), S1. https://doi.org/10.1016/0013-4694(91)90210-u

Reilly, S., Wake, M., Bavin, E. L., Prior, M., Williams, J., Bretherton, L., … Ukoumunne, O. C. (2007). Predicting Language at 2 Years of Age: A Prospective Community Study. *Pediatrics*, *120*(6), e1441–e1449. https://doi.org/10.1542/peds.2007-0045

Rescorla, L., Mirak, J., & Singh, L. (2000). Vocabulary growth in late talkers: lexical development from 2;0 to 3;0. *Journal of Child Language*, *27*(2), 293–311. https://doi.org/10.1017/S030500090000413X

Roberts, M. Y., & Kaiser, A. P. (2015). Early Intervention for Toddlers With Language Delays: A Randomized Controlled Trial. *PEDIATRICS*, *135*(4), 686–693. https://doi.org/10.1542/peds.2014-2134

Sanyaolu, A., Okorie, C., Qi, X., Locke, J., & Rehman, S. (2019). Childhood and Adolescent Obesity in the United States: A Public Health Concern. *Global Pediatric Health*, *6*, 2333794X1989130. https://doi.org/10.1177/2333794X19891305

Smith, V., & Mirenda, P. (2009). Concurrent and predictive validity of the MacArthur-Bates Communicative Development Inventory for children with autism. In *International Meeting for Autism Research* (p. May 7-9). Chicago.

Stothard, S. E., Snowling, M. J., Bishop, D. V. M., Chipchase, B. B., & Kaplan, C. A. (1998). Language-Impaired Preschoolers: A Follow-Up Into Adolescence. *Journal of Speech, Language, and Hearing Research*, *41*(2), 407–418. https://doi.org/10.1044/jslhr.4102.407

Sy, S. R., Gottfried, A. W., & Gottfried, A. E. (2013). A Transactional Model of Parental Involvement and Children's Achievement from Early Childhood through Adolescence. *Parenting*, *13*(2), 133–152. https://doi.org/10.1080/15295192.2012.709155

Tardif, T., & Fletcher, P. (2008). *Chinese communicative development inventories: user's guide and manual*. Beijing: Peking University Medical Press.

Thal, D. J., Desjardin, J. L., & Eisenberg, L. S. (2007). Validity of the MacArthur-Bates Communicative Development Inventories for measuring language abilities in children with cochlear implants. *American Journal of Speech-Language Pathology*. https://doi.org/10.1044/1058-0360(2007/007)

Thompson, E. C., Krizman, J., White-Schwoch, T., Nicol, T., Estabrook, R., & Kraus, N. (2019). Neurophysiological, linguistic, and cognitive predictors of children's ability to perceive speech in noise. *Developmental Cognitive Neuroscience*, *39*(August 2018), 100672. https://doi.org/10.1016/j.dcn.2019.100672

White-Schwoch, T., Woodruff Carr, K., Thompson, E. C., Anderson, S., Nicol, T., Bradlow, A. R., … Kraus, N. (2015). Auditory Processing in Noise: A Preschool Biomarker for Literacy. *PLOS Biology*, *13*(7), e1002196. https://doi.org/10.1371/journal.pbio.1002196

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), 420–423. https://doi.org/10.1038/nn1872

Table 1. Summary of results of predictive models for two- and three-way classifications for different outcomes of the MCDI. The non-neural features include sex, birth weight and gestational age. The added neural features include the 69 EEG features (see **SI**) and the gap between the neural data recording and language outcome assessment. AUC = Area under the receiver-operating curve. Mean predictive values, their standard deviations and above-chance significance in bootstrap-permutation test are shown (*$p < .05$, **$p < .01$, ***$p < .001$)

Predictor	Feature selection	Measure	MCDI Outcomes			
			Early Gestures	Later Gestures	Vocabulary Comprehension	Vocabulary Production
Three-way prediction (<25th vs. 25th – 75th vs. >75th percentile)	Non-neural only	Accuracy,%	58±6.2***	66.7±5.66***	54.2±6.39***	62.2±6***
		Sensitivity,%	58.2±11.5*	77.8±7.04 ns	57.6±11.6 ns	59.1±12**
		Specificity, %	75±7.64 ns	67±9.25***	7.6±8.83 ns	78.2±6.67 ns
		AUC	.67±.070**	.74±.06***	.64±.073*	.67±.075**
	Neural features added	Accuracy, %	73.5±4.39***	77.5±4.17***	72.8±4.38***	74.2±4.46***
		Sensitivity, %	76.8±8.02***	87±5.13***	74.7±8.01***	69.4±9.41***
		Specificity, %	86.7±4.83**	72.2±7.36***	83.3±5.55**	88.6±4.45**
		AUC	.88±.041***	.89±.033***	.85±.047***	.85±.05***
Two-way prediction (<25th vs. >25th percentile)	Non-neural features only	Accuracy, %	69.6±5.76*	72.9±5.51***	66.3±6.14*	72.2±5.6*
		Sensitivity, %	51±12.3*	74.9±7.65*	47.7±12.5*	51.9±13.3**
		Specificity, %	79.6±6.97 ns	69.7±8.57**	76.6±7.8 ns	81.1±6.15 ns
		AUC	.69±.069**	.75±.059***	.65±.072*	.71±.071**
	Neural features added	Accuracy, %	84.4±3.47***	79.7±3.96***	82.2±3.76***	84±3.65***
		Sensitivity, %	72.6±8.26***	80.9±5.87**	70.9±8.17**	67.2±9.33***
		Specificity, %	90.7±3.7**	77.7±6.77***	88.4±4.37**	91.6±3.68**
		AUC	.92±.031***	.91±.028***	.90±.035***	.89±.039***

Figure legends

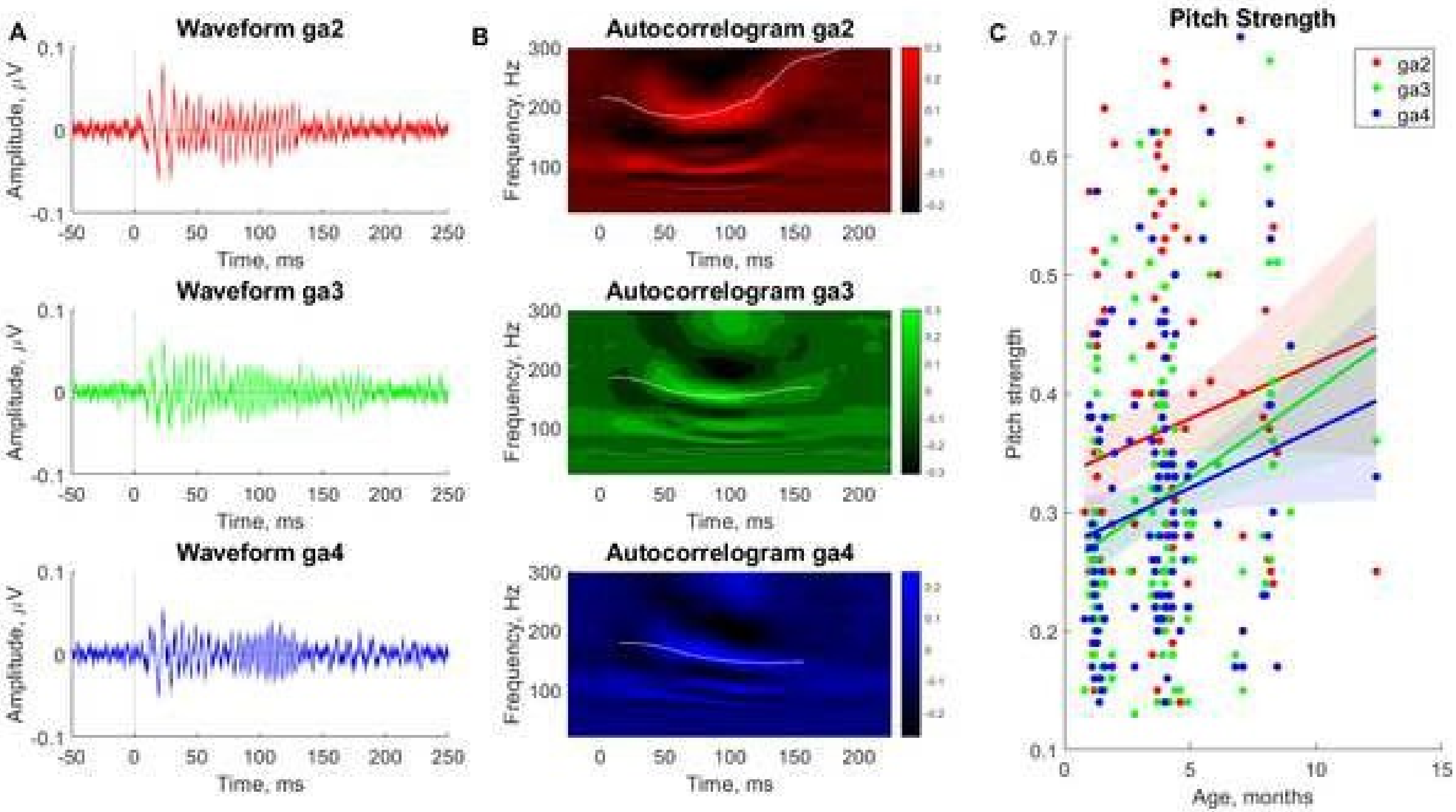
Figure 1. The waveforms (**A**) and autocorrelograms (**B**) of the short-latency responses to the three speech stimuli. The pitch contours on autocorrelograms overlap with the pitch of the corresponding auditory stimuli (white line). The rapid maturation of short-latency responses is illustrated by the growth of the Pitch Strength (maximal autocorrelation), which is closely related to the age of individual members of our group of participants (**C**)

Figure 2. The waveforms of the long-latency (cortical) response show a prominent positive peak typical for this age group (**A**). This response is stable across the ages in our experimental group, as illustrated by the lack of growth of its signal-to-noise ratio (**B**).

Figure 3. Neural measures combined with non-neural measures outperformed non-neural measures alone in predicting language outcomes as indicated by significant increases in AUC for both 2-way (**A**) and 3way (**B**) language outcome responses. The *p*-value is given in the upper right corner of each graph for the critical comparison of non-neural vs. non-neural + neural models. The outcomes are, from top to bottom: Early Gestures, Late Gestures, Vocabulary Comprehension and Vocabulary Production in the MacArthurBates Communicative Development Inventories – Cantonese version. The non-neural models consist of sex, birth weight and gestational age (red). The neural models include additionally 69 neural features (see **SI**) and the gap between the neural data recording and language outcome assessment (green). All models' AUC values are significantly higher than noise as measured by permuted-label models (grey) (*p*-values not shown). Solid lines are the medians and the dotted lines are 5th and 95th percentiles of the corresponding distributions

Figure 4. Prediction of language outcomes as measured by AUC is not affected either by the age of neural data recording (**A**) or by the length of the gap between the neural data recording and language outcome assessment (**B**). In both cases the two models were constructed from the two subsets of data as determined by a median split of the original dataset. The median age of our experimental group was 3.8 months, while the median EEG-language assessment gap was 8.6 months. The models for younger and smaller age-gap groups are shown in pink, while the models for older and larger age-gap groups are shown in green. *p*values are given in the upper right corner of each graph for the critical comparisons. All models significantly outperformed their permutations (*p*-values not shown). The solid lines are the medians and the dotted lines are the 5th and 95th percentiles of the corresponding distributions.

Figure 1



A
Waveform ga2
Waveform ga3
Waveform ga4
Amplitude, μV
Time, ms
B
Autocorrelogram ga2
Autocorrelogram ga3
Autocorrelogram ga4
Frequency, Hz
Time, ms
C
Pitch Strength
ga2
ga3
ga4
Pitch strength
Age, months

Figure 2

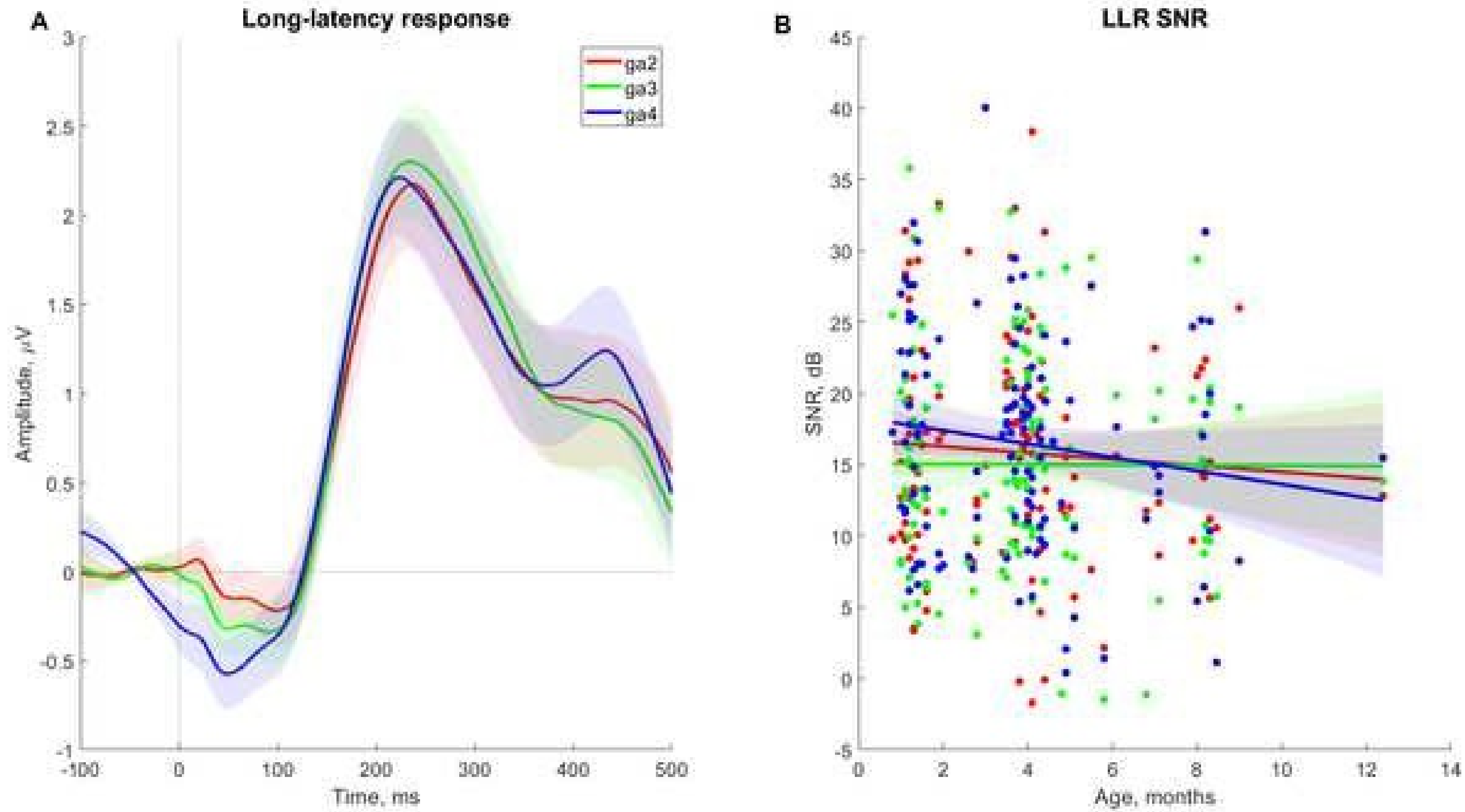
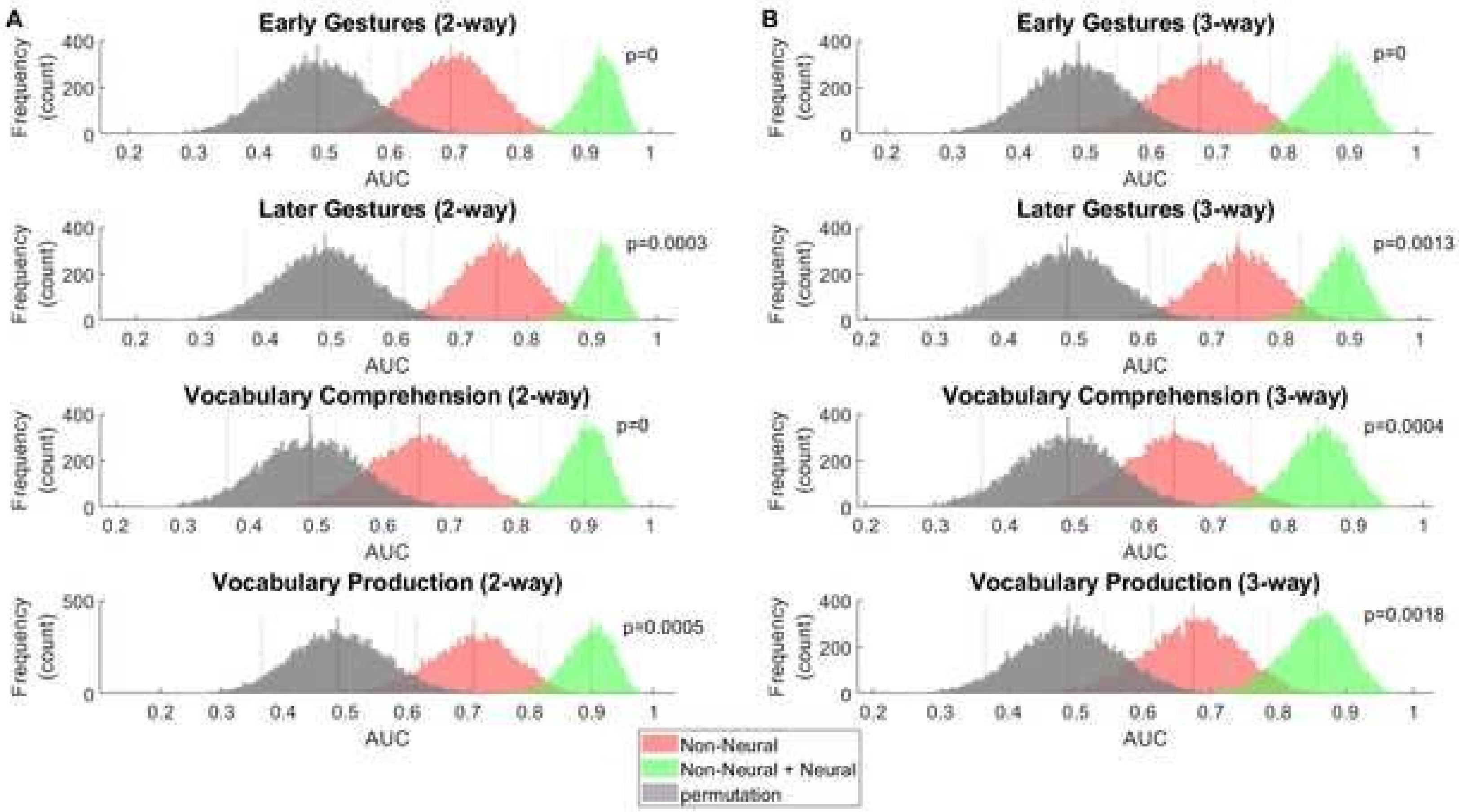
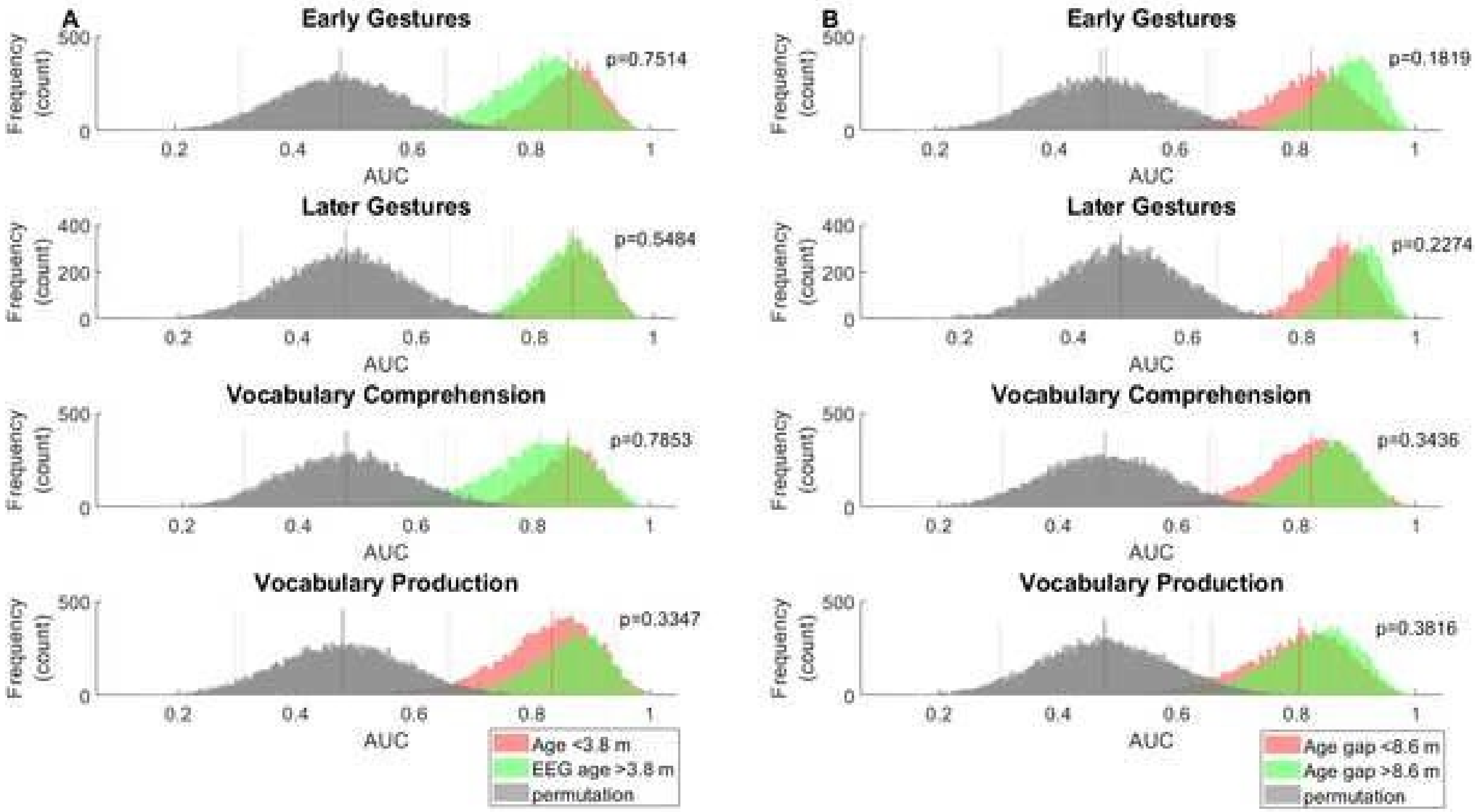

A
Long-latency response
ga2
ga3
ga4
Amplitude, μV
Time, ms
B
LLR SNR
SNR, dB
Age, months

Figure 3



A
Early Gestures (2-way)
p=0
Later Gestures (2-way)
p=0.0003
Vocabulary Comprehension (2-way)
p=0
Vocabulary Production (2-way)
p=0.0005
B
Early Gestures (3-way)
p=0
Later Gestures (3-way)
p=0.0013
Vocabulary Comprehension (3-way)
p=0.0004
Vocabulary Production (3-way)
p=0.0018
Frequency (count)
AUC
Non-Neural
Non-Neural + Neural
permutation

Figure 4



A
Early Gestures
p=0.7514
Later Gestures
p=0.5484
Vocabulary Comprehension
p=0.7853
Vocabulary Production
p=0.3347
Age <3.8 m
EEG age >3.8 m
permutation
B
Early Gestures
p=0.1819
Later Gestures
p=0.2274
Vocabulary Comprehension
p=0.3436
Vocabulary Production
p=0.3816
Age gap <8.6 m
Age gap >8.6 m
permutation
Frequency (count)
AUC