

Introduction

Several decades of research on infants has established that children are born with a universal capacity to perceive phonetic differences across all languages, but by 12 months of age, infant phoneme perception narrows such that native-language differences are recognized more accurately while non-native phoneme discrimination declines (Kuhl et al. 2006; Kuhl et al. 1992; Werker et al. 1981; Werker & Tees 1984). Moreover, researchers have demonstrated that speech perception skills in infancy are a reliable predictor of future language abilities. For example, Tsao et al. 2004 showed that individual variation in infants' speech discrimination skills at 6 months of age-predicted their language abilities at three future points in time, 13 months, 16 months, and 24 months [REF].

We used MEG to record the auditory evoked MMN, whose time course is taken as an electrophysiological index for early speech-comprehension processes (Pulvermuller F. 2006). MMN is a neuro-physiological index of the detection of a change in the acoustic input that can be elicited in the absence of focused attention (for a review, see Näätänen et al., 2007). It arises in the oddball paradigm when listeners are confronted with a series of stimuli, some of which are frequently presented (standards) and some infrequently (deviants). Relative to the response evoked by standards, around 200 ms after stimulus onset, deviants evoke a more pronounced response – negative in EEG. This is labeled Mismatch Negativity, MMN.

The rationale behind the oddball paradigm combines what is known about early infant language learning [REFS] and categorical perception of speech sounds [REFS]. It is well known that at birth, infants are able to detect differences between speech sounds used in the world's languages. However, within the first year of life, this capacity is altered by language experience resulting in a narrowing of phonetic perceptual ability facilitating native language learning (Cheour et al. 1998; P. K Kuhl et al. 1992; P K Kuhl et al. 2006; Conboy et al. 2008; Rivera-Gaxiola, Silva-Pereyra, and Kuhl 2005; Sundara, Polka, and Genesee 2006). Along with this, we also know that information like speech sounds is typically perceived as discrete categories as opposed to gradual instances when there is a change in some variable along a continuum (Ingram and Jusczyk 1999). In our paradigm, we use categorical tokens of syllables /ba/ and /wa/ as deviants along with a standard CV syllable conveying a consonantal onset with pre-voicing period (VOT) halfway between VOT durations in deviants. As such, the standard stimulus is equally likely to be perceived as either deviant CV in a sampling of English speaking. Thus, we reasoned that infants at 2-months will hear three distinct CV stimulus and present with typical (adult-like) mismatch activity. However, in light of early phonetic learning, by 6-months, infants are less likely to hear three distinct syllables, and thus show differential or degraded auditory mismatch activity due to the emergence of categorical phonological processing.

Materials & Methods

Subjects

Seventy-five (34 male) typically developing (TD) English-learning infants between 2-4 months-age were recruited for the study. At the time of the initial visitation, children were on average 79 ± 30.2 (M \pm STD) days old, and based on parental reporting, were born at full-term between 39 - 42 weeks gestational age; on average weighing between 8 ± 2.0 lbs. and head sized 40 ± 1.6 cm in circumference. Of the total, 39 infants (17 male, average age 192 ± 6.5 days old, average head circumference 44 ± 1.8 cm) returned for a follow-up MEG examination at six-months-age. Children included here had no history of ear infections or other hearing difficulties. Ethical approval was obtained from the University of Washington Human Subjects Division. As per the Declaration of Helsinki, the parents or legal guardians of all children provided informed written before participation.

Stimuli

Stimuli consisted of prototype tokens consonant-vowel (CV) syllabic speech segments (/ba/ and /wa/) synthesized to uniquely convey a phonological VOT contrast between plosive-labial [b] and liquid [w] consonants (Pisoni, Carrell & Gans, 1983). Briefly, the VOT contrast consisted of consonant segments with a 20-ms period of low amplitude pre-voicing ([b] 15ms, and [w] 45ms) followed by variable first and second formant transitions (F1, F2), where F1 started at 234 Hz and rose linearly to 769 Hz, and F2 started at 616 Hz and rose linearly to a steady-state value of 1232 Hz. For the steady-state vowel segments, F3 – F5 energy levels were constant at 2862 Hz, 3600 Hz, and 3850 Hz. (cf. fig 1). Speech sounds were delivered as digital audio files (325ms in duration) at 65 dB SL via a free-field speaker situated 1 m center-front of the subject, with digital-to-analog conversion and amplification (Tucker-Davis Technologies).

Stimulation

To evoke the auditory evoked mismatch, we used a multi-deviant auditory oddball paradigm with variable stimulus onset asynchrony (Refs). Stimulus sound files response containing CV segments were presented in pseudo-randomized arrangements of consecutive oddball stimuli train. The standard stimulus was a CV token consisting of a consonantal segment with a 30ms or mid-point VOT yielding that was perceptually ambiguous, which means that the standard stimulus was categorically perceived as /ba/ or /wa/ at chance in a randomly chosen sample of native American-English speaking adults (supp). For each subject, oddball stimuli were arranged, such that: (a) the same deviant never occurred consecutively and (b) a deviant was always separated by at least three occurrences of the standard. Each stimulus was separated by onset asynchrony of 1.0 ± 0.5 s [Expyfun]. At the onset of each session, 80 consecutive trials of the standard stimulus were presented to establish a full auditory memory trace. Each recording session lasted approximately 15 minutes in length, yielding 680 trials of auditory evoked response field data per subject.

MEG data acquisition

As permitted, all data acquisition, processing procedures were carried out per recommendations in []. Each individual infant was prepared for MEG testing outside the MSR while a research assistant entertained them. A 3D position monitoring digitization system (Polhemus, Colchester, VT) was used to record the locations of head position indicator (HPI) coils, cardinal (nasion, left/right preauricular) anatomical landmarks, and additional points (> 100) covering the scalp. We used an infant seat made for use in the MEG scanner to record brain activity in participants. The child's head was centered and positioned as high as possible relative to the MEG dewar, using foam cushions and padding to adjust the position for recording. All infants were awake and alert during recordings. A female research assistant entertained the infant using silent toys, accompanied by a silent video in the background throughout the recording session to maintain arousal. Neuromagnetic MEG signals were recorded with a 306-channel whole-head MEG system (VectorView™, Elekta Neuromag Oy, Helsinki, Finland) with active shielding inside a single-layered MSR at the Institute for Learning & Brain Sciences, University of Washington, Seattle. Neuromagnetic data was sampled at 1 kHz with a pass-band of .01 to 300 Hz. During MEG recordings, signals from the HPI coils were used to continuously track the child's head position relative to the fixed position of the MEG sensors.

Data reduction and analysis

MEG data signal processing was carried out using MNE-python (Gramfort et al. 2013, 2014). To suppress signal artifacts from external sources, the data were processed using temporal signal space separation (Taulu, Kajola, and Simola 2004). The continuous HPI data was used to compensate for subjects' head movements using the initial head position as the target MEG coordinate frame. Signal space projection (Haamalainen 1995) was used to suppress the cardiac signal from the MEG data by estimating the spatial structure of the QRS complex from a maximally responsive gradiometer sensor for each subject. Single trials of neuromagnetic data were averaged to create ERF datasets for each oddball stimuli using 0.7s (0.1 s pre-stimulus) epochs. To adjust for DC offset in evoked data, the mean amplitude across sensors in the pre-stimulus interval was subtracted from sensor values in the post-stimulus period.

Language measures

Language skills were assessed using the MacArthur-Bates Communicative Development Inventories (CDIs), Words, and Sentences assessment [REF]. The words and sentences form are designed to measure language production in children from 16 to 30 months of age. This form divides language production into two parts. Part 1 contains a 680-word vocabulary production checklist. Part 2 includes five sections designed to assess morphological and syntactic development. We used three of these sections in this study: vocabulary size, sentence complexity, and mean length of the longest three utterances (MLU). For each child included in the study, the CDI survey was completed online by the parent or caregiver survey starting at 15 months of age, i.e., the earliest time point the words and sentences form can be administered,

with follow-ups at three months intervals through 30 months of age, for each completed survey families received \$10 for their effort.

Results

Table 1 describes the demographic features for N=72 (33 male) infants with longitudinal language outcome measurements between 18-30 months, and electrophysiological data including n=24 (11 male) longitudinal MEG exams at 2- and 6-months age. As expected, AUC classifier scores for auditory evoked responses to CV stimuli conveying a consonantal VOT contrast between labial-plosive [b] to liquid [w] were differentially distributed over developmental age.

Table.1-Sample demographics & descriptive summary

	Group	Age(days)	Weight(oz)	Head size(cm)	SES	Maternal Edu(yrs)	Paternal Edu
<i>N</i>	six	24	24	24	24	24	24
	four	17	17	17	17	17	17
	two	31	31	31	31	31	31
<i>Mean</i>	six	191.75	116.63	43.75	52.54	16.5	16.42
	four	124.88	127	42.29	46	15.76	15.47
	two	59.45	119.26	39.48	53.03	16.52	16.13
<i>Standard deviation</i>	six	7.16	11.62	1.36	12.1	1.91	2.7
	four	12.33	11.87	1.4	15.64	2.44	3.43
	two	6.01	13.35	1.09	11.84	1.86	2.83
<i>25th percentile</i>	six	188	110.5	43	42.75	16	15
	four	115	117	42	40.5	14	12
	two	54.5	111	39	43	16	14.5
<i>50th percentile</i>	six	189.5	115	44	56.5	17	16.5
	four	122	133	43	43	16	15
	two	59	115	39	57	17	16
<i>75th percentile</i>	six	197	124	44.25	61.25	17.25	18
	four	135	137	43	61	17	18
	two	64	126.5	40	61.25	18	18

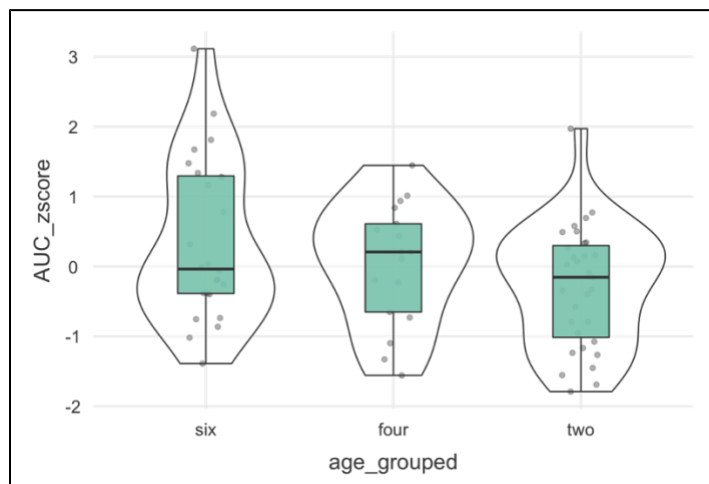


Figure.1-Logit classifier

The distributions of AUC scores from early measurements shown in figure 1 indicates an upward trend in mismatch reliability as represented here by an increase in classifier reliability scores between 2- and 6-months early infant phonetic learning irrespective of VOT contrast.

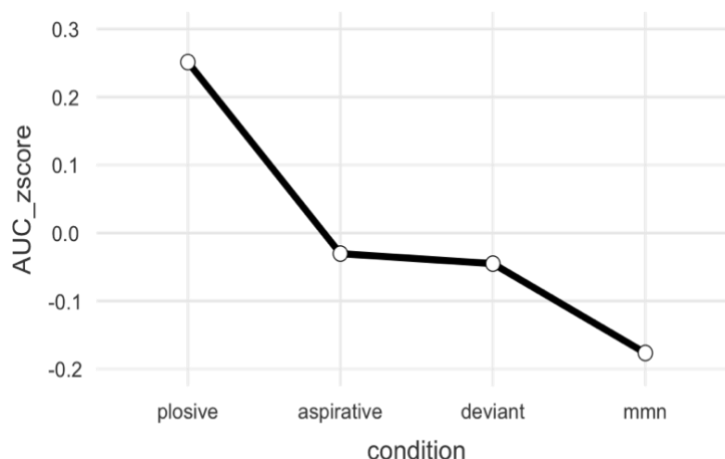
The upward trend in AUC scores with age was evaluated with mixed linear modeling to estimate the additive effect of infant age and oddball conditioning accounting for confounding influence of SES, gender, birth weight and head circumference. Mixed linear model fit by REML confirmed that classifier reliability was effected by oddball condition ($F_{(3,1362)}=20.03$, $p<.001$). Post-hoc repeated measures comparison of cell means revealed that in comparison to the shorter VOT [b] (higher acoustic energy) CV stimulus as deviant; AUC scores for all other classifier conditions were significantly attenuated ($p'_{\text{bonferroni}} <.001$).

Table.2-Post-hoc comparisons

logit score comparison						
A	B	Difference	t	df	$p_{\text{bonferroni}}$	
plosive	- MMN	0.43	0.06	7.52	1362	<.001**
plosive	- liquid	0.28	0.06	4.95	1362	<.001**
plosive	- deviant	0.3	0.06	5.21	1362	<.001**
liquid	- MMN	0.15	0.06	2.57	1362	0.062
liquid	- deviant	0.01	0.06	0.26	1362	1
deviant	- MMN	0.13	0.06	2.31	1362	0.126

Figure 2 shows the EMMs for AUC scores from logit classifier fit to mean ERF amplitude across magnetic gradiometers in the *a priori* MMN time window following VOT contrasts: (plosive:[æ] \rightarrow [ba]); liquid: [æ] \rightarrow [wa]; MMN: [æ] \rightarrow CV_{ERF}; Deviants: [ba] \rightarrow [wa]. On average AUC scores for the plosive contrast (0.52 ± 0.05 ; $M \pm SD$) were 0.43 standardized deviation scores larger than scores in the MMN condition (0.50 ± 0.04). AUC scores for scalp level mean ERF amplitude following a deviant CV oddball stimulus (irrespective of VOT length) yielded smaller AUC values as compared to the case where the deviant stimulus encoded a plosive (high acoustic energy) VOT contrast. In fact, in relative terms both VOT contrast AUC scores were significantly larger than MMN condition combining ERFs for VOT contrasts (see table 2 for stats).

Figure.2-Logit classifier EMMs



LMM ANOVA results also confirmed that classifier reliability was modulated by age. Fixed effects omnibus testing revealed a significant interaction term between condition and age ($F_{(3,1362)}=21.03$, $p<.001$). Examination of model parameter estimates indicated that differential patterning of auditory mismatch activity with age. However, with the difference with

age was driven by a significant reduction in AUC scores ($\beta=-0.01\pm9.8e^{-04}$, $t_{1362}=-7.77$, $p<.001$) between deviant ERF amplitudes and mismatch activity combining VOT contrasts in the deviant position (i.e. acoustic mismatch).

Table.2-Fixed effects parameter estimates

Effect	Estimate	SE	95% Confidence Interval		df	t	p
			Lower	Upper			
(Intercept)	-1.41e-4	0.08	-0.15	0.15	68	0	0.999
plosive - liquid	0.28	0.06	0.17	0.39	1362	4.95	<.001**
liquid - deviant	0.01	0.06	-0.1	0.13	1362	0.26	0.796
deviant - MMN	0.13	0.06	0.02	0.24	1362	2.31	0.021*
boy - (girl)	0	0.15	-0.31	0.3	68	-0.02	0.983
ses	-1.44e-4	0.01	-0.01	0.01	68	-0.02	0.981
age	3.06E-04	0	0	0	68	0.23	0.818
plosive - liquid * age	0	9.81E-04	0	5.69E-04	1362	-1.38	0.168
liquid - deviant * age	0.01	9.81E-04	0	0.01	1362	5.24	<.001**
deviant - MMN * age	-0.01	9.81E-04	-0.01	-0.01	1362	-7.77	<.001**

Discussion

Logistic classifier cross-validation routine used to evaluate differential auditory evoked activity in response to tokens of CV syllabic

stimuli conveying different consonantal VOT contrasts. Considering the MMN is a known electrophysiological index of phonological contrast processing in the adult AC, then the AUC score for classifier reliability to distinguish ERF time series features in response to different oddball events represents differential auditory processing in infant brain.

- Cheour, Marie, Rita Ceponiene, Anne Lehtokoski, Aavo Luuk, Juri Allik, Kimmo Ahlo, and Risto Naatanen. 1998. "Development of Language-Specific Phoneme Representations in the Infant Brain." *Nature Neuroscience* 1 (5): 351–53. <https://doi.org/10.1038/1561>.
- Conboy, B T, M Rivera-Gaxiola, J F Silva-Pereyra, and P K Kuhl. 2008. "Event-Related Potential Studies of Early Language Processing at the Phoneme, Word, and Sentence Levels." In *Early Language Development: Bridging Brain and Behaviour, Trends in Language Acquisition Research Series*, edited by A D Friederici and G Thierry, 5:24–64. Amsterdam/The Netherlands: John Benjamins.
- Gramfort, A, M Luessi, E Larson, D A Engemann, D Strohmeier, C Brodbeck, R Goj, et al. 2013. "MEG and EEG Data Analysis with MNE-Python." *Front Neurosci* 7: 267. <https://doi.org/10.3389/fnins.2013.00267>.
- Gramfort, A, M Luessi, E Larson, D A Engemann, D Strohmeier, C Brodbeck, L Parkkonen, and M S Haamaalaainen. 2014. "MNE Software for Processing MEG and EEG Data." *Neuroimage* 86: 446–60. <https://doi.org/10.1016/j.neuroimage.2013.10.027>.
- Haamaalaainen, M S. 1995. "Functional Localization Based on Measurements with a Whole-Head Magnetometer System." *Brain Topogr* 7 (4): 283–89. <https://www.ncbi.nlm.nih.gov/pubmed/7577326>.
- Ingram, David, and Peter Jusczyk. 1999. *The Discovery of Spoken Language. Language*. Vol. 75. <https://doi.org/10.2307/417480>.
- Kuhl, P. K, K. A Williams, F Lacerda, K. N Stevens, and B Lindblom. 1992. "Linguistic Experience Alters Phonetic Perception in Infants by 6 Months of Age." *Science (New York, N.Y.)* 255 (5044): 606–8. <https://doi.org/10.1126/science.1736364>.
- Kuhl, P., K. Williams, F Lacerda, K. Stevens, and B Lindblom. 1992. "Linguistic Experience Alters Phonetic Perception in Infants by 6 Months of Age." *Science* 255 (5044): 606–8. <https://doi.org/10.1126/science.1736364>.
- Kuhl, P K, E Stevens, A Hayashi, T Deguchi, S Kiritani, and P Iverson. 2006. "Infants Show a Facilitation Effect for Native Language Phonetic Perception between 6 and 12 Months." *Developmental Science* 9: F13–21.
- Kuhl, Patricia K., Erica Stevens, Akiko Hayashi, Toshisada Deguchi, Shigeru Kiritani, and Paul Iverson. 2006. "Infants Show a Facilitation Effect for Native Language Phonetic Perception between 6 and 12 Months." *Developmental Science* 9 (2). <https://doi.org/10.1111/j.1467-7687.2006.00468.x>.
- Rivera-Gaxiola, Maritza, Juan Silva-Pereyra, and Patricia K. Kuhl. 2005. "Brain Potentials to Native and Non-Native Speech Contrasts in 7- and 11-Month-Old American Infants." *Developmental Science* 8 (2): 162–72. <https://doi.org/10.1111/j.1467-7687.2005.00403.x>.

- Sundara, Megha, Linda Polka, and Fred Genesee. 2006. "Language-Experience Facilitates Discrimination of /d-Th/ in Monolingual and Bilingual Acquisition of English." *Cognition* 100 (2): 369–88. <https://doi.org/10.1016/j.cognition.2005.04.007>.
- Taulu, Samu, Matti Kajola, and Juha Simola. 2004. "The Signal Space Separation Method." *ArXiv physics.in*.
- Werker, J. F., J. H. Gilbert, K. Humphrey, and R. C. Tees. 1981. "Developmental Aspects of Cross-Language Speech Perception." *Child Dev* 52 (1): 349–55. <https://doi.org/10.1111/j.1467-8624.1981.tb03051.x>.
- Werker, J. F., and R. C. Tees. 1983. "Developmental Changes across Childhood in the Perception of Non-Native Speech Sounds." *Canadian Journal of Psychology* 37 (2): 278–86. <https://doi.org/10.1037/h0080725>.