

ARTICLE

Twenty-month-olds categorically discriminate similar sounding vowels regardless of vocabulary level, an event related potentials (ERP) study

Ao CHEN

School of Psychology, Beijing Language and Culture University, China
Institute for Language Sciences, Utrecht University, the Netherlands
Email: Irischen71@hotmail.com

(Received 23 August 2022; revised 06 February 2023; accepted 24 April 2023)

Abstract

The current study investigated whether vocabulary relates to phonetic categorization at neural level in early childhood. Electroencephalogram (EEG) responses were collected from 53 Dutch 20-month-old children in a passive oddball paradigm, in which they were presented with two nonwords “*giep*” [ɣip] and “*gip*” [ɣɪp] that were contrasted solely by the vowel. In the multiple-speaker condition, both nonwords were produced by twelve different speakers; while, in the single-speaker condition, one single token of each word was used as stimuli. Infant positive mismatch responses (p-MMR) were elicited in both conditions without significant amplitude differences. When the infants were median split based on vocabulary level, the large and small vocabulary groups showed comparable p-MMR amplitudes yet different scalp distribution in both conditions. These results suggest successful phonetic categorization of native similar sounding vowels at 20 months, and a close relationship between speech categorization and vocabulary development.

Keywords: Phonetic categorization; vocabulary; mismatch response

Introduction

Infants are surrounded by rich and constantly varying speech input. For example, the same word can drastically differ acoustically when produced in different sentences, by different speakers, with different emotions, or for different purposes. One formidable task that infants face when learning language is to extract the invariable linguistic content from varying acoustical input, a task that adults accomplish with ease (Tuninetti et al., 2017). Yet despite the various approaches adopted, when and how children succeed in phonetic categorization (i.e., to perceive acoustically discriminable tokens as functionally equivalent) has been a central yet still not fully understood question for language and cognitive science (Choi & Shukla, 2021; Crinnion et al., 2020; Mulak et al., 2017; Schatz et al., 2021; Swingley & Alarcon, 2018).

With electroencephalogram (EEG), the current study tested 20-month-old toddlers’ phonetic categorization of similar sounding native vowels, and aimed to find neural

signatures under categorical discrimination. The toddlers' brain responses were collected when they passively listened to a minimal pair of consonant-vowel-consonant (CVC) nonwords (i.e., the two nonwords solely differed in the vowel). Furthermore, this study investigated whether vocabulary size could be a contributing factor in speaker normalization at the neural level.

Normalizing speaker variability is an important aspect of phonetic categorization. The same phoneme or word can drastically differ acoustically when produced by different speakers, due to the differences in size of the vocal tract, pattern of vocal fold vibration, and palate shape, as well as accent and social status (Harnsberger et al., 2010; Johnson & Sjerps, 2018), yet adults readily recognize the linguistic content of speech forms in the face of this variability. Normalizing speaker variability is a formidable task for young language learners when abstracting phonetic categories. Infants and young children are sensitive to both linguistic (i.e., vowel) and indexical (i.e., speaker) change of isolated vowels (Mulak et al., 2017). Two studies indicated that by five years, children still had more difficulty recognizing words when these were produced by multiple speakers than by one single speaker (Goldinger et al., 1991; Ryalls & Pisoni, 1997). Speaker variability is particularly challenging for infants' discrimination of acoustically similar vowels, since these overlap in terms of the first (F1) and second (F2) formants, the primary acoustic determinant of vowels, and individual speakers differ in terms of absolute F1 and F2 values (Feldman, Griffiths, et al., 2013; Johnson & Sjerps, 2018).

Different hypotheses have been proposed with regard to how infants learn to disambiguate overlapping phonetic categories. The distributional learning account hypothesizes that infants form phonetic categories by grouping the sounds that cluster in the perceptual space (Maye et al., 2002). However, such account has been questioned seeing that across languages, F1/F2 distribution of vowels in infant directed speech is insufficient to reliably delineate vowel boundaries (Cristia & Seidl, 2014; Englund & Behne, 2005; Kuhl et al., 1997; Martin et al., 2015; McMurray et al., 2013), and some evidence reveals that infants are sensitive to phonetic details that are too fine-grained and, thus, may hinder categorization (Schatz et al., 2021). Other studies have argued that word knowledge may guide infants to dissociate overlapping vowel categories. Given that infants learn to segment and recognize word forms in the same period as they learn to establish native phonological categories, knowledge of word forms (with or without semantic content) may scaffold discrimination of overlapping phonetic categories among preverbal infants (Feldman, Myers, et al., 2013; Yeung & Werker, 2009). For example, the distribution of /ɛ/ and /æ/ overlap acoustically in English across different speakers; however, if the vowel in "cat" always has a higher F2 and lower F1 than the vowel in "bed" in the speech input, the lexical context may serve as an index of the different distribution of the acoustic characteristics of vowels (Swingle, 2009; Swingle & Alarcon, 2018). Furthermore, the lexical restructuring model hypothesizes that vocabulary expansion, or the pairing of sound forms to meanings, is a driving force for refining phonological representation. For instance, knowing that /jɪp/ pairs with one object and /jɪp/ with another will help infants to segment the words into phonemes and to understand that /i/ and /ɪ/ are different vowels that distinguish word meaning (Metsala & Walley, 1998; Walley et al., 2003). Therefore, word templates, or the context in which particular phonemes occur, may serve as a helpful cue for young children to disambiguate phonetic categories that cannot be readily distinguished by acoustic information.

So far, however, the relationship between phonetic categorization and vocabulary development has not been tested directly. Previous studies have shown that behavioral and neural discrimination of native phonemes in the first year of life correlate with both

concurrent and future vocabulary (Garcia-Sierra et al., 2011; Kuhl et al., 2008; Singh, 2019). Yet it is in the second year of life that infants' vocabulary starts to quickly expand (Goldfield & Reznick, 1990; Nazzi & Bertoncini, 2003), and as hypothesized by the lexical restructure model (Metsala & Walley, 1998; Walley et al., 2003), the vocabulary effect on phonetic categorization is expected to be most evident at the early stage of a vocabulary spurt. At the early stage of vocabulary development, substantial individual variability exists (Chen et al., 2017; Fenson et al., 1994). One toddler may already know hundreds of words while another knows only a dozen. If word knowledge indeed relates to phonetic categorization, then those toddlers with a large vocabulary should be more capable of categorizing speech sounds than those with a small vocabulary.

To test the concurrent association between word knowledge and phonetic categorization, the current study tested Dutch 20-month-old toddlers on their neural discrimination of acoustically similar native /ɪ/ and /i/ contrast embedded in CVC nonwords (i.e., *giep* [ɣɪp] and *gip* [ɣip]), and examined whether vocabulary level relates to categorical discrimination of the nonwords. In Dutch, both /i/ and /ɪ/ are front unrounded vowels, and they cannot be unequivocally distinguished by duration (Adank et al., 2004). By 20 months, toddlers were still not able to fully accurately encode minimally different vowels in both familiar and novel words (Mani et al., 2008; Nazzi, 2005). Thus, speaker normalization of these vowels was expected to be challenging for toddlers. Such immature and developing vowel representation, together with the quick vocabulary expansion at this age, should make the mutual influence between vowel categorization and word knowledge (if any) evident. In the current study, the infants' vocabulary was measured by the Dutch version McArthur-Bates Communicative Development Inventory (N-CDI) (Zink & Lejaegere, 2002). If word knowledge facilitates vowel categorization, when speaker variation was introduced, the toddlers with a large vocabulary should exhibit stronger neural discrimination of the nonwords than those with a small vocabulary.

Neural discrimination of the stimuli was tested with mismatch responses (MMRs). For adults, mismatch negativity (MMN; (Peltola et al., 2003; Saloranta et al., 2020) is a component of auditory event-related potentials (ERP). MMN can be elicited using a passive oddball paradigm, in which listeners are presented with a stream of repeating 'standard' sounds conforming to a certain regularity punctuated occasionally by 'deviant' sounds, dissimilar in some relevant dimension from the standards. If the brain detects the change from standard to deviant, then on the difference waveform obtained by subtracting the response to the standard from that to the deviant, the MMN is visible as a negative peak between 100 and 300 ms from deviant onset in adults (Näätänen et al., 2007). In addition to physical difference between the standard and deviant, MMN can also be elicited by violation of abstract patterns (Xiao et al., 2018), and listeners are able to extract similarity within the variable standards and deviants and detect dissimilarity across the two types. Previous research has shown that for adults, while the brain is sensitive to speaker difference (Tuninetti et al., 2017), it is able to separate the phonologically contrastive vowel categories in the face of speaker variability (Shestakova et al., 2002, 2004).

For infants and young children, unlike adults, MMN does not yield consistent temporal and spatial characteristics. A commonly observed developmental pattern is the shift of mismatch response polarity – namely, that at an early stage of development, infants' and children's mismatch responses tend to exhibit a late positivity (positive mismatch response, p-MMR) rather than the adult MMN (He et al., 2009; Morr et al., 2002). As infants grow older, the polarity of the mismatch response shifts to negative, and gradually approximates the adult MMN (Bishop et al., 2011; Morr et al., 2002). The shift

from p-MMR to MMN has been found to occur earlier for acoustically salient (i.e., when the stimuli showed larger physical difference) than non-salient stimuli (Cheng et al., 2015; He et al., 2009; Lee et al., 2012). In addition, several studies have demonstrated co-occurring p-MMR and MMN among infants (Leppänen et al., 1997; Morr et al., 2002; Yu et al., 2019), and, thus, the p-MMR and MMN are likely to be generated by difference neural processes.

Speech categorization continues to develop through adolescence (McMurray et al., 2018). For consonants, some studies found evidence for successful speaker normalization at the neural level in early infancy (Dehaene-Lambertz & Baillet, 1998; Dehaene-Lambertz & Pena, 2001; van Leeuwen et al., 2006). Yet, given the high degree of acoustic overlap between the vowels, it remains inconclusive whether young children are able to normalize speaker variability when discriminating similar sounding vowels. Furthermore, the neural signature underlying vowel categorization among infants is largely unknown. A number of previous studies made use of one single standard and one single deviant as stimuli to test neural discrimination of vowels, and results are inconsistent with regards to the MMR polarity and scalp distribution (Čeponien et al., 2003; Cheour-Luhtanen et al., 1995; Lee et al., 2012; Marklund et al., 2019; Shafer et al., 2012). For example, MMN (i.e., NEGATIVE mismatch response) was reported for native synthesized vowels among Finnish newborns (Cheour-Luhtanen et al., 1995) as well as preschool Finnish children (Čeponien et al., 2003). English learning infants, on the other hand, showed a p-MMR to the acoustically non-salient native /ɪ/ and /e/ contrast up to 30 months, and afterwards adult-like MMN started to emerge (Shafer et al., 2011, 2012). A further study demonstrated that among the young children one early MMR (160-360 ms) and one late MMR (400-600 ms) co-occurred, and both shifted gradually from positive to negative as the infants grew older (Yu et al., 2019). For Swedish infants, the native /i/-/e/ contrast elicited p-MMR among 4- to 8-month-olds (Marklund et al., 2019). For Mandarin infants, the mismatch response elicited by the large change from /a/ to /u/ (i.e., differed in both roundedness and backness) shifted from positive to negative at 6 months, while the mismatch response elicited by the small change from /a/ to /i/ (i.e., both are front and unrounded) stayed positive in preschool years (Cheng et al., 2015; Lee et al., 2012). In addition, Shafer et al. (2012) suggested that, in addition to maturity, attention may modulate the polarity of the mismatch responses among infants. Importantly, seeing that only one single token of each vowel was used as stimuli in these studies, whether and when the infant brain is able to discriminate between vowel categories remains unknown. It should be noted however, 2-3-month-old infants showed mismatch response to a non-native vowel contrast after being exposed to a bimodal distribution of 900 acoustically different tokens of these vowels for a few minutes (Wanrooij et al., 2014), and by three months, the infant brain was able to encode consonants in the manner and place of articulation, which were elementary units robust to variability (Gennari et al., 2021). To sum up, the development of vowel elicited MMR seems to be language and stimuli specific, and although previous research hinted at early categorization of vowels at the neural level, direct experimental evidence is lacking.

To fill in the void, the current study tested 20-month-old toddlers' phonetic categorization in two separate conditions. In the experimental condition, both the nonwords *gip* [ɣɪp] and *gip* [ɣɪp] were produced by 12 different speakers; thus, only if the infant brain was able to normalize speaker variation and regard the 12 different tokens of each nonword as the "same", would they be able to neurally discriminate the two nonwords. In the control condition, one single token of each vowel produced by one speaker was used as stimuli, and neural discrimination would succeed if the toddlers could detect the

acoustical difference between two stimuli. Four hypotheses were proposed. First, if the toddlers were able to normalize speaker variability, a significant mismatch response (either positive or negative) should be observed in the multiple-speaker condition. Second, if variability overburdened vowel discrimination, the mismatch response should be attenuated in the multiple-speaker compared to the single-token condition. Third, if categorical perception of vowels was well established by 20 months, the mismatch response should exhibit an adult-like negative polarity. Fourth, if vocabulary level related to neural discrimination of the vowels, the toddlers with a large vocabulary should exhibit either a more evident MMR or differ from those with a small vocabulary in terms of MMR polarity.

Materials and methods

Participants

Fifty-nine 20-month-old infants participated in the current study. Six infants were excluded from analysis due to crying or excessive head movements. The remaining 53 participants (29 girls, age range 19 months 1 day to 20 months 23 days) had a mean (SD) age of 601 (13) days, and all were all full-term healthy infants from monolingual Dutch families, and no hearing or neurological disorders were reported by the parents. All the parents gave written consent for participating in the experiment. The experiment was approved by Utrecht University Faculty Ethics Assessment Committee Humanities.

Materials

Twelve female native Dutch speakers (mean age = 21 years, SD = 2.5 years) were recruited to produce the stimuli. They were first visually familiarized with a list of printed CVC nonword minimal pairs, where the two nonwords in a pair were solely distinguished by the vowel (e.g., *tos* [tos] and *toes* [tus], *nief* [nif] and *nif* [nif]). They could spend as much time as they needed to get familiar with these nonwords. The target nonwords for the current experiment, *giep* [ɣip] and *gip* [ɣip], formed one of these pairs. The rest of the nonwords were for a word learning experiment which together with the current experiment formed parts of a larger project examining early phonology and word development. Next, the participants were asked to produce all the nonwords in carrier sentences *ik zei niet X maar ik zei Y* as well as *ik zei niet Y maar ik zei X* (meaning I did not say X but I said Y, and vice versa), where X and Y were one of the minimal pairs. They were told to speak the sentences in a way as if they were talking to a toddler. The speakers were recorded with a Sennheiser ME-64 microphone and a DAT recorder TASCAM DA-40 in a sound attenuated room.

For each speaker, one well-realized token of [ɣip] and one token of [ɣip] were cut off from the recording for further manipulation in PRAAT (Boersma & Weenink, 2011). All the tokens had a falling f₀ contour. The duration of the tokens was manipulated to have a mean of 344 ms (SD = 9.5 ms, range 323–361 ms) and the intensity was scaled to 70dB. These manipulated [ɣip]s and [ɣip]s were used stimuli in the current experiment. All the twelve tokens of each nonword were presented in the multiple-speaker condition, while one token of each nonword from one speaker was presented in the single-speaker condition. Duration and mean f₀ were measured for the vowel part (i.e., /i/ and /ɪ/), and F1, F2 and F3 values were measured at the temporal midpoint of the steady part of the vowels. The [ɣip] and [ɣip] used in the single-speaker condition had an F1 of 348 Hz and

485 Hz, an F2 of 2827 Hz and 2361 Hz, and an F3 of 3432 Hz and 3100 Hz respectively. The acoustical characteristics of the vowels in the stimuli were consistent with those reported in previous studies (Adank et al., 2004). Multiple native Dutch speakers listened to the stimuli and reported the stimuli to be natural, and all were able to identify all the stimuli as [ɪp] or [ɣɪp] correctly. The acoustic measurements of the stimuli are listed in Table 1 in the supplementary materials. The oscillograms of the stimuli can be found in Figure 2 in the supplementary materials.

We used CVC nonwords rather than isolated vowels as stimuli because /i/ can occur as a reduced format of *hij* (meaning he) in colloquial Dutch, such as in the sentence *wat doet ie* (meaning *what does he do*), whereas /i/ alone can never be a word. Using nonwords as stimuli precluded lexical status from being a confounding factor. Although the Dutch /i/ and /ɪ/ may be considered to contrast in duration besides F1/F2, acoustical analysis has shown that duration does not distinguish these vowels sufficiently (Adank et al., 2004; Swingley, 2019). Dutch, however, does have long and short vowels, such as in *maan* [ma:n] (meaning moon) and *man* [mɑn] (meaning man), and both Dutch adults and infants were found to be sensitive to long vowels being mispronounced as short ones but not vice versa (Chládková et al., 2015; Dietrich et al., 2007). Therefore, to prevent duration from being a confounding factor, the vowels in [ɪp] and [ɣɪp] were not manipulated to contrast in duration.

In both conditions, [ɪp] was the standard and [ɣɪp] was the deviant. It should be acknowledged that according to previous studies on vowel perception, to detect a change from a less to a more peripheral vowel was easier than the other way around (Kuhl, 1991; Polka & Bohn, 2011) – hence, the assignment of standard and deviant may have an influence on the mismatch response. Yet it was not our purpose to investigate asymmetry in neural detection of the vowel change. The effect of word knowledge could be better interpreted with a consistent assignment of standard and deviant across the participants, since any relationship found between the MMR and vocabulary cannot be due to the confounding factor of standard/deviant assignment.

The infants' vocabularies were measured by the Dutch version McArthur-Bates Communicative Development Inventory: Words and Sentence (N-CDI) (Zink & Lejaegere, 2002), which includes a total of 702 words, divided into 22 different semantic categories.

Procedure

A passive oddball paradigm was adopted. Infants' brain responses were recorded in two blocks: a multiple-speaker block followed by a single-speaker block. As the focus of the current study was phonetic categorization, the multiple-speaker block always preceded the single-speaker condition.

Each block comprised 600 trials, of which 480 (80%) were standards and 120 (20%) deviants, and each trial was composed of a single sound token. Each block began with 10 trials of the standard, after which standards and deviants were presented continuously in a pseudo-random order with the constraint that deviants were separated by at least two standards. The inter-trial interval was randomly varied between 320 ms and 400 ms.

The EEG was recorded in a sound-attenuated room at Utrecht University. The infant participants sat on their caregivers' laps during the experiment. Infant friendly silent animated videos were played on the computer screen, and parents were instructed not to talk during the experiment. Toys were placed on the table in front of the infant, with

which they could play if they wanted to. The distance between the participant's eyes and the screen was ~1 m and the experimental stimuli were presented at 70dB SPL (measured at where the infant sat) through two audio speakers on each side of the screen. EEG was recorded with a Biosemi system from a 32-channel cap with Ag/AgCL electrodes according to the 10–20 International System of Electrode Placement. EEG was recorded at a sampling rate of 2048 Hz. During online recording, a 5th order Bessel filter for optimal pulse response was used by Biosemi system, and the -3 dB frequency of the filter is placed at 1/4 of the (fixed) sample frequency.

The parents filled in the N-CDI at home online. For each word, they were asked to indicate by mouse clicking whether their child “understands but does not produce yet”, “understands and produces”, or “does not understand and does not produce yet”. The raw scores were automatically generated with locally developed software. The N-CDI was filled in either before or after the experiment, and there was an average of eight days (SD = 9 days) between the experiment and the filling of the N-CDI.

EEG processing

The EEG data were analysed offline using EEGLAB toolbox (version 13.1.1b in Matlab 2015b, (Delorme & Makeig, 2004). The raw recordings were filtered between 0.3–20 Hz, and down-sampled to 250 Hz. The continuous recordings were re-referenced to the average of all electrodes and were segmented into 700ms epochs from the 100ms before the onset (baseline) to 600ms after the stimulus onset. Continuous bad channels were visually inspected and Spline-interpolated. 27 participants had channels interpolated, and on average 0.96 (SD = 0.88) channels was interpolated. Trials having an amplitude larger than ± 150 microvolts were removed. The standards immediately after a deviant were excluded from analysis. The remaining artefact-free trials were averaged to obtain the ERPs for each infant. Infants who had more than 25 artefact-free deviant trials were included in the final analysis, and all the infants met this criterion. Individual waveforms were averaged to obtain the grand averaged waveform. In the multiple-speaker condition, mean (SD) number of accepted standard and deviant trials were 271 (34) and 91 (11) respectively, and in the single-speaker condition 265 (45) and 88 (15) respectively.

Statistical analysis

A two-step analysis was adopted. First, to test whether the standard and the deviant ERPs differed significantly for any time window, with all the participants collapsed, for each condition, point-by-point *t* tests were performed with the standard and deviant ERPs (i.e., ERPs of individual participant) for all the points between 200 and 600ms after the stimulus onset for F3, Fz, and F4 separately. If for at least one electrode, the standard and deviant ERPs significantly differed for at least six consecutive time points (i.e., 24 ms with the sampling rate being 250 Hz), the difference was considered as meaningful (Chen et al., 2018; Cheng et al., 2015). Subsequently, an MMR peak was identified on the grand average. Second, for amplitude measures, individual MMR peak latencies were identified in the 100ms window (50ms before and after) surrounding the grand average peak, and individual MMR peak amplitudes were calculated as the mean amplitude in the 40 ms (20 ms before and after) window surrounding the corresponding individual peak. The mean amplitude of the standard and deviant ERPs was calculated for the same time window as for the MMR. As MMR is most evident for frontal and

central locations, the frontal (F3, Fz, F4) and central electrodes (C3, Cz, C4) were included in the analysis. To investigate whether the MMRs were more frontally or centrally distributed, a repeated measures ANOVAs with type (standard or deviant), site (front or central electrodes), lateralization (left, middle, or right) and condition (multiple- and single-condition) being the within-subject variables were conducted with individual peak amplitudes at the six electrodes. As the significant interaction between type and site confirms the frontal distribution of the MMRs (see Results), further analyses were conducted with the MMR amplitudes at F3, Fz, and F4 to investigate the effect of condition.

To investigate the effect of vocabulary size, the participants were median split into a high comprehension group (HC, $N = 26$) and a low comprehension group (LC, $N = 27$) based on the total comprehension score (i.e., the sum of the words checked by parents as “understands but does not produce yet” and “understands and produces”) of the N-CDI. Seeing that the p-MMRs were most evident at the frontal sites (see analyses below), p-MMR amplitudes were examined with mixed effect ANOVAs with lateralization (left F3, middle Fz, or right F4), and condition (multiple- and single-speaker) being the within-subject variable and vocabulary level (HC and LC) being the between-subject variable. The same analysis was also conducted with productive scores, which can be found in [supplementary materials](#).

Results

All participants

Table 1 lists the time points where the standard and deviant ERPs differed significantly as shown by the point-by-point t tests. Figure 1 plots the standard ERPs, the deviant ERPs and the difference waves in the multiple- and single-speaker conditions averaged across all the participants.

As can be seen from Table 1 and Figure 1, a significant p-MMR was observed for each condition between 300 ms and 400 ms after the stimulus onset based on the point-by-point t tests, and these P-MMRs were most visible at F3. Therefore, grand average p-MMR peaks were identified between 300ms and 400ms at F3 for each condition, and the grand average had a peak latency of 336ms and 356ms in the multiple- and single-speaker condition, respectively. When individual peak latencies were averaged, the p-MMRs had a mean (SD) peak latency of 332 ms (25 ms) in the multiple-speaker condition and 359 ms (27 ms) in the single-speaker condition, and the latency difference between the two conditions was significant, $t(52) = 5.28$, $p < .001$, partial $\eta^2 = .35$. Table 2 in the supplementary materials lists the mean peak amplitudes at F3, Fz, F4, C3, Cz, and C4 for the multiple- and single-speaker condition. Figure 2 plots the p-MMR topography at

Table 1. Time windows (ms after stimulus onset) where the standard and the deviant ERPs differed significantly for the multiple-speaker (MS) and single-speaker (SS) conditions. (+) indicates that the deviant ERP was more positive than the standard ERP, and (-) indicates that the deviant ERP was more negative than the standard ERP, and – indicates lack of significant difference

	F3	Fz	F4
MS	292-404(+)	324-368(+)	–
SS	336-404(+)	588-596(+)	–

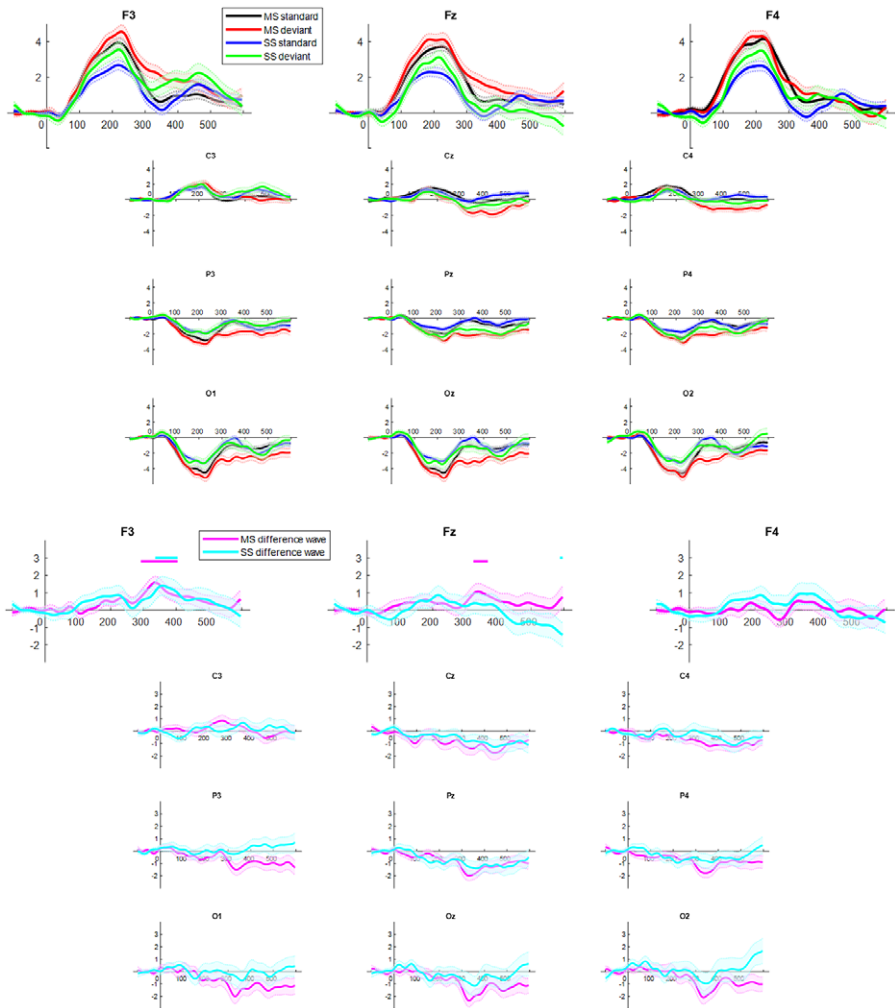


Figure 1. Standard [ɣip] and deviant [ɣip] ERPs (upper panel) and difference waves (lower panel) averaged across all the participants. MS = multiple-speaker condition, SS = single-speaker condition. Lines above x axis indicate time windows where the difference waves were significantly larger than 0 (i.e., positive MMR), and lines below x axis indicate time windows where the difference waves were significantly smaller than 0 (i.e., negative MMR). Shaded area represents standard errors.

the corresponding grand average peak latency in the multiple- and single-speaker condition.

Repeated measures ANOVAs with type (standard or deviant), site (front or central electrodes), lateralization (left, middle, or right) and condition (multiple- and single-condition) being the within-subject variables found significant main effect of site (frontal or central), $F(1, 52) = 46.93, p < .001, \text{partial } \eta^2 = .47$, and lateralization (left, middle, or right) $F(2, 104) = 13.89, p < .001, \text{partial } \eta^2 = .21$, showing that the ERPs were larger at frontal than at central electrodes, and they were more positive at the left than the right

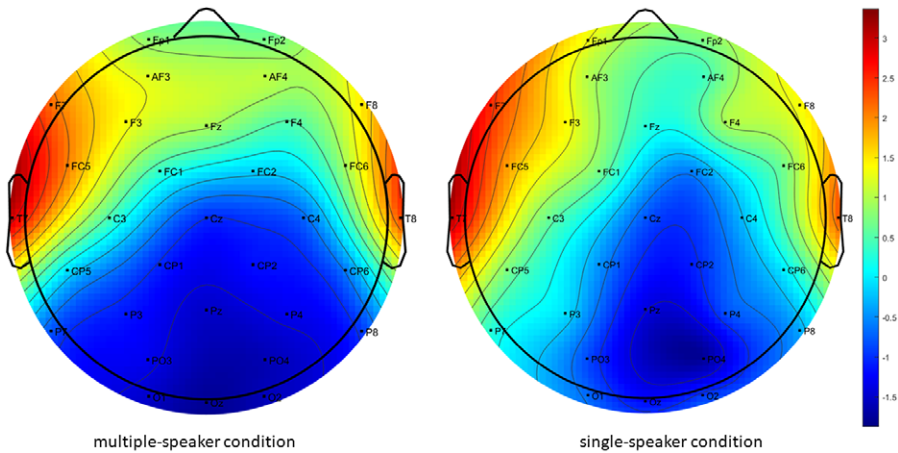


Figure 2. p-MMR topography at peak latency in the multiple-speaker (336ms) and single-speaker (356ms) conditions.

electrodes. The main effect was not significant for either condition $F(1, 52) = 1.48, p = .23$, partial $\eta^2 = .03$ or type $F(1, 52) = 1.85, p = .18$, partial $\eta^2 = .03$ turned out to be significant. The interaction between conditions and site was significant $F(1, 52) = 13.80, p < .001$, partial $\eta^2 = 0.21$, and so was the interaction between type and site $F(1, 52) = 20.39, p < .001$, partial $\eta^2 = 0.28$. Together with Figure 1, it can be seen that the differences between standard and deviant ERPs were most evident at frontal electrodes, and this difference is much attenuated at the central electrodes, and overall, the ERPs are left lateralized. Therefore, further analysis was conducted with the MMR amplitudes at F3, Fz, and F4. There was a significant main effect of lateralization, $F(2, 104) = 5.57, p = .005$, partial $\eta^2 = 0.10$ while the main effect of conditions $F(1, 52) = 0.61, p = .44$, partial $\eta^2 = 0.01$, and the interaction between the two factors $F(2, 104) = 0.55, p = 0.58$, partial $\eta^2 = 0.01$ were not significant. Post-hoc Bonferroni corrected pairwise comparison showed that the p-MMR at F3 was significantly larger than that at Fz and F4 ($p < .05$).

These results showed that in both the single- and the multiple-speaker condition, frontally distributed p-MMRs were elicited, and they had comparable peak amplitudes across the conditions. With regard to latency, the p-MMR had an earlier peak latency in the multiple-speaker than in the single-speaker condition.

Comprehensive vocabulary groups

There are in total 702 words listed in the N-CDI. The participants had a mean comprehension raw score of 318 (SD = 126, range 50-571) and a mean production raw score of 133 (SD = 133, range 3-408). A median (270) split was applied to the comprehension raw scores (i.e., total number of words checked by parents as “understands but does not produce yet” and “understands and produces”) of CDI, and those infants who had a score higher than the median were considered as high comprehenders (HC, N=26) and the rest low comprehenders (LC, N=27). The HC group had a mean comprehension score of 419 (SD = 84) and a mean production score of 188 (SD = 125), while those for the LC group were 219 (SD = 69) and 80 (SD = 63) respectively. The same median split was also

Table 2. The time windows (ms, after stimulus onset) where the standard and deviant ERP differed significantly and mean of individual p-MMR peak latencies (ms, after stimulus onset) in the multiple-speaker (MS) and single-speaker (SS) conditions, separated for HC (high comprehenders) and LC (low comprehenders). (+) indicates that the deviant ERP was more positive than the standard ERP, and (-) indicates that the deviant ERP was more negative than the standard ERP, and – indicates lack of significant difference

condition	vocabulary level	F3	Fz	F4	mean (SD) p-MMR peak latencies at F3
MS	HC	224-248(+) 296-488(+)	292-444(+)	–	332 (22)
	LC	324-352(+)	–	–	333 (26)
SS	HC	–	–	200-292(+) 316-432(+)	357 (27)
	LC	320-424(+)	–	468-504(-)	364 (26)

applied to the production raw scores. The high producers produced a mean of 221 words ($SD = 95$), and the low producers produced a mean of 48 words ($SD = 30$).

Table 2 lists the time windows where the standard and deviant ERPs differed significantly and the mean of individual p-MMR peak latencies for the HC and LC group separately. As can be seen, point-by-point t tests identified p-MMRs for both groups in both conditions. As the difference between standard and deviant ERPs was most frequently observed at F3 (except for HC in the single-speaker condition), grand average p-MMR peak latencies were identified between 300ms and 400ms after the stimulus onset at F3. The LC group had a grand average peak latency of 336 ms and 356 ms and the HC group 340 ms and 364 ms in the multiple- and single-speaker condition respectively. Table 3a and Table 3b in the supplementary materials list the individual amplitudes of the standard, deviant ERPs, and the p-MMR. Figure 3 plots the difference waves of the HC and LC groups in the multiple- and single-speaker condition. The standard and deviant ERPs of the two groups can be found in the supplementary materials. Figure 4 plots the p-MMR topography of the high and low comprehenders at the corresponding grand average peak latencies in the multiple- and single-speaker condition.

For individual p-MMR peak amplitudes (i.e., the mean amplitude in the 40 ms window surrounding individual p-MMR peaks), electrode (F3, Fz, F4) * condition (single- or multiple-speaker) * comprehension group (HC, LC) mixed effect ANOVAs found no significant main effect of condition, $F(1, 51) = .58, p = .45$, partial $\eta^2 = .01$, or vocabulary group, $F(1, 51) = 1.83, p = .18$, partial $\eta^2 = .04$. The main effect of electrode was significant, $F(2, 102) = 5.95, p = .004$, partial $\eta^2 = .11$. Bonferroni corrected post hoc pairwise comparison showed that overall, F3 had a larger amplitude than Fz and F4. The interaction between electrode and vocabulary level was significant, $F(2, 102) = 6.78, p = .002$, partial $\eta^2 = .12$. None of the other interactions was significant. Bonferroni corrected post hoc analysis found that for the LC group, p-MMR had a larger amplitude at F3 than Fz and F4 while there was no significant difference between the electrodes for the HC group.

With regard to peak latencies, condition showed a significant main effect, $F(1, 51) = 39.51, p < .001$, partial $\eta^2 = 0.44$, and main effect of vocabulary level was not significant, $F(1, 51) = 2.46, p = .12$, partial $\eta^2 = 0.05$. The interaction between the two was not significant, $F(1, 51) = 0.24, p = .72$, partial $\eta^2 = 0.01$. Similar to the analysis with all

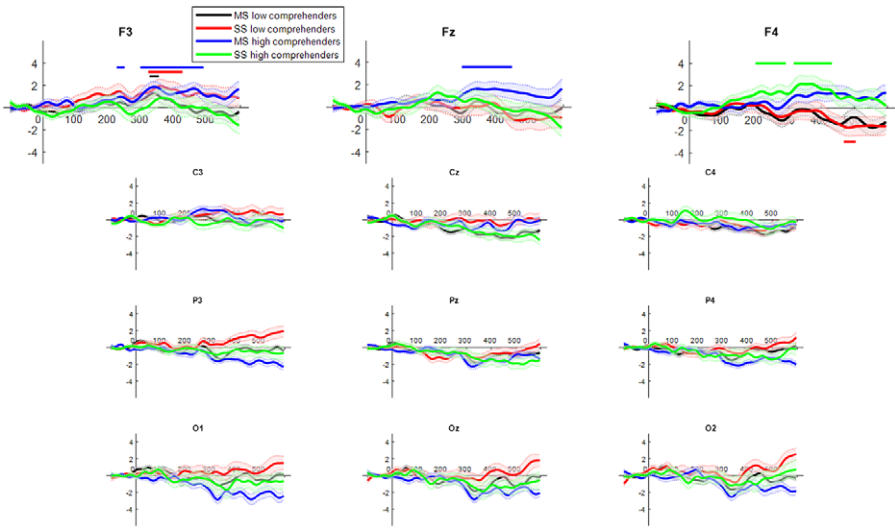


Figure 3. The difference waves of the high and low comprehenders in the multiple-speaker (MS) and single-speaker (SS) conditions.

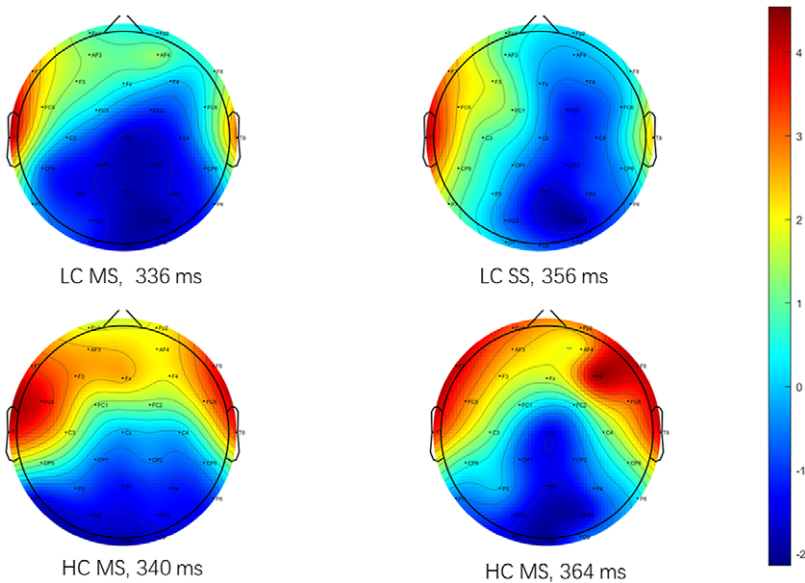


Figure 4. p-MMR topographies of the high comprehenders (HC) and low comprehenders (LC) at peak latency in the multiple-speaker (MS) and single-speaker (SS) conditions.

participants, both the HC and LC groups had an earlier peak latency in the multiple- than in the single-speaker condition.

Figure 3 together with the statistical analyses showed that the p-MMR amplitudes were comparable between the high and low comprehenders, and both groups had an earlier

p-MMR peak latency in the multiple- than in the single-speaker condition. Nevertheless, the two groups showed different p-MMR scalp distributions: while the p-MMR was symmetrically distributed for the HC group, it was left lateralized for the LC group.

General Discussion

In the current study, we investigated 20-month-old infants' neural detection of vowel change with and without presence of speaker variability. One token of each nonword [ɣip] and [yip] from each speaker constituted the multi-speaker stimuli, and in the single-speaker condition, one token of each from the same speaker were used as stimuli. For both conditions, the [ɣip] occasionally embedded in the stream of [yip] elicited significant positive mismatch response (p-MMR), with the peak latency being 336 ms (significant window: 292-404 ms) and 356 ms (significant window: 336-404 ms) for the multiple- and single-token condition respectively. The presence of speaker variability did not attenuate p-MMR, as no significant p-MMR amplitude difference was found across the conditions. To investigate how vocabulary influenced categorization, the infants were median split into high and low comprehenders/producers based on their raw score of the Dutch version CDI (words and sentences). In terms of p-MMR peak amplitude, the high and low groups were comparable for both the single- and multiple-speaker conditions. Group specific patterns, however, were also observed. First, the low group showed a left-lateralized distribution while the high group showed a symmetrical scalp distribution of p-MMR for both the conditions. Second, in the multiple-speaker condition, the high group showed a more sustained p-MMR than the low group.

One crucial finding of the current study is that infants as young as 20 months were already able to neurally discriminate the acoustically similar vowels in the face of speaker variability. The p-MMR in the multiple-speaker condition was not attenuated compared to the single-speaker condition, where the brain only had to detect the acoustical difference. Furthermore, the p-MMR was not influenced by vocabulary. Previous studies have shown neural discrimination of native vowels in early infancy (Čeponien et al., 2003; Cheour-Luhtanen et al., 1995; Lee et al., 2012; Marklund et al., 2019; Shafer et al., 2012), yet using only one single token of each category as stimuli, it remains unclear whether infants were responding to the stimuli acoustically or categorically. By including the multiple-speaker condition, the current study demonstrated at 20 months, children have already acquired perceptual constancy of similar sounding vowels at the neural level, and the infant brain was able to disregard speaker information when discriminating the vowels. Thus, the abstract representation of native phonetic categories becomes well established early in life.

At this age, however, the mismatch response still showed a positive polarity, which was consistent with previous studies (Lee et al., 2012; Shafer et al., 2010), and a shift to adult-like MMN is expected to occur at a later age. Some previous studies have shown that among infants and toddlers, the p-MMR can co-occur with a negativity at a late time window, yet whether such a late negativity reflects reorientation of attention or it is equivalent to adult MMN remains debated (Kushnerenko et al., 2013; Shafer et al., 2010; Yu et al., 2019). In the current study, no significant late negativity was observed, except for the low comprehenders in the single-speaker condition. Since we only tested one age group, it is difficult to ascertain whether the late negativity will emerge at a later age, and whether the p-MMR latency and amplitude will change with age (Shafer et al., 2000, 2010). Seeing that the current study made use of different stimuli, SOA, and procedure

compared to the previous studies, no conclusion could be drawn with regard to the mechanisms underlying the emerging late negativity. It is important to compare the mismatch responses elicited by different stimuli at the same age as well as those elicited by the same stimuli but at different ages. Another developmental pattern that needs further investigation in the future is the scalp distribution of the p-MMR. In the current study, as can be seen on the topographic maps, the p-MMRs are largest at lateral site, while previous studies often reported frontal distribution of the MMRs (Morr et al., 2002; Shafer et al., 2011; Yu et al., 2019). Other studies found that while p-MMR was left lateralized, MMN emerged from frontal central sites (Shafer et al., 2010). It might be that for the toddlers tested in the current study, the MMN was emerging, and the overlap between MMN and p-MMR created the lateral distribution. But again, the developmental change of MMR topography needs to be investigated by including more age groups.

As the vowels were presented in nonwords, the successful neural discrimination cannot be attributed to familiarity with the stimuli, nor to semantic knowledge. In addition, as the multiple-speaker condition always preceded the single-speaker condition, the infants had no chance to first establish targets from invariant stimuli in the single-speaker condition and then map the variable tokens to the targets. Speaker variation has been found to facilitate word learning among infants (Rost & McMurray, 2010), and it has been hypothesized that variation along linguistically irrelevant dimensions helps infants identify the phonologically relevant acoustical dimensions and consequently facilitates sound-meaning pairing. The current study, however, did not teach infants new words but tested what they had already learned by the time of the experiment. Evidently, these similar sounding nonwords had already been well contrasted in the infant brain by the time of the experiment. Therefore, the p-MMR in the multiple-speaker condition likely reflects existing categorical representation of the vowels.

Inconsistent with our hypothesis, the results do not support the enhancement effect of vocabulary on categorical discrimination of the vowels, given that the HC and LC group's p-MMRs were comparable in terms of both amplitude and peak latency. Seeing that both the lexical restructuring model (LRM, (Metsala & Walley, 1998) and the lexical-distribution models (Feldman, Myers, et al., 2013; Swingley, 2009) argue that word knowledge facilitates phonetic learning, as words may provide specific contextual cues for discovering and disambiguating the phonetic categories, we speculate that the mutual influence between word learning and phonetic categorization might be more evident at an earlier age. The infants in the current study understood 318 words on average, and the number of words known by the LC group might have been sufficient to support discriminating the similar sounding vowels. In particular, perhaps knowing word forms without meaning might have sufficiently facilitated learning of phonetic categories (Carbajal et al., 2021). Whether it is word forms or word meanings that associate to phonetic category needs further investigation.

Nevertheless, group specific characteristics were observed. First, for the multiple-speaker condition, the high group showed a more symmetrical scalp distributions of p-MMR across the two conditions while for the low group, the p-MMR scalp distribution was left-lateralized. Second, the high group showed a more sustained p-MMR than the low group in the multiple-speaker condition. MMR has been found to last longer for more salient than less salient phonemic contrasts (Cheng et al., 2015; Lee et al., 2012), implying that for the same contrast, a longer lasting MMR of the high group might reflect a stronger response to the contrast. Our results seem to suggest that a large vocabulary might be helpful when infants discriminated between variable vowel categories. Interestingly, although different cortical lateralization has been found for vowel versus speaker

processing, as well as for within- versus cross-category phoneme discrimination (Maurer et al., 2003; Poeppel et al., 2004; Shestakova et al., 2004; Sittiprapaporn et al., 2005; Xi et al., 2010), in the current study, for neither group was the MMR scalp distribution different across conditions. Instead, regardless of whether speaker variation was present, the p-MMR differed across vocabulary levels. Therefore, it seems that vocabulary had a general effect, although not necessarily a facilitative one, on speech sound discrimination, independent of the presence of variability. Several previous studies have shown that speech perception in the first year of life, operationalized mainly as discrimination accuracy of single exemplars of different native speech phonetic categories, correlated with later vocabulary ability (Garcia-Sierra et al., 2011; Kuhl et al., 2008; Singh, 2019; Tsao et al., 2004). The current study found that, concurrently, different neural networks might be at play for categorical discrimination of vowels for toddlers with different vocabulary level. For now, it is hard to ascertain whether it is age or vocabulary that serves as the main driving force for the co-occurring development in word learning and phonological categorization between 18 and 24 months. Biological maturation, accumulative language input, development of cognitive skills, or a combination of all may all play a role (Werker et al., 2002; Werker & Curtin, 2005), and it is worth the effort for future studies to focus on this period and identify the crucial factors bolstering phonology and vocabulary development.

It should be acknowledged that two alternative explanations of the findings cannot be ruled out at this moment. One possibility is that the toddlers responded to the acoustic distance between the standards and the deviants rather than vowel categories. The acoustic distance might be smaller among the different tokens of the same than the different vowel type, and the toddler brain might have made use of such acoustic cues in discrimination. It would be informative if future studies can manipulate the stimuli in a way that the acoustic distance between variable tokens is equivalent both within and across the standard and deviant types. Second, unlike adult MMN, which is widely agreed to indicate discrimination (Bartha-Doering et al., 2015; Fu & Monahan, 2021; Hisagi et al., 2010, 2015; Trainor et al., 2003), infant p-MMR may simply reflect recovery from refractoriness. To eliminate the potentially different obligatory responses elicited by different sounds, it is important for future studies to include equiprobable presentation of the standard and deviant stimuli or to flip the standard and deviant for comparison. In addition, seeing that the current study tested normalization of intra-speaker variability, how toddlers normalize inter-speaker variability is unknown. It would be informative for future studies to investigate how vowel categorization may differ when different types of variation are present.

To conclude, children as young as 20 months were able to neurally discriminate similar sounding vowels successfully, regardless whether speaker variability was incorporated and regardless whether they had a small or large vocabulary. Nevertheless, the brain seemed to be activated differently as the result of vocabulary, leading to different scalp distribution of the mismatch responses.

Acknowledgements. This study was funded by the Dutch Research Council with the project number 275-89-034 and supported by The National Social Science Fund of China with the project number 17CYY041, and by Science Foundation of Beijing Language and Culture University (the Fundamental Research Funds for the Central Universities, Approval Number 21PT01). I thank Lisanne Geurts and Charlotte Koevoets for their help with data collection. I thank Maartje de Klerk, Frank Wijnen, Likan Zhan, Ruohan Chang, Varghese Peter and the Babylab in Utrecht Institute of Linguistics for their very insightful discussions. I thank all the parents and babies who took part in the study.

Supplementary material. The supplementary material for this article can be found at <http://doi.org/10.1017/S0305000923000351>.

Competing interest. The author(s) declare none.

References

- Adank, P., van Hout, R., & Smits, R. (2004). An acoustic description of the vowels of Northern and Southern Standard Dutch. *The Journal of the Acoustical Society of America*, *116*(3), 1729–1738. <https://doi.org/10.1121/1.1779271>
- Bartha-Doering, L., Deuster, D., Giordano, V., am Zehnhoff-Dinnesen, A., & Dobel, C. (2015). A systematic review of the mismatch negativity as an index for auditory sensory memory: From basic research to clinical and developmental perspectives. *Psychophysiology*, *52*(9), 1115–1130. <https://doi.org/10.1111/psyp.12459>
- Bishop, D., Hardiman, M. J., & Barry, J. G. (2011). Is auditory discrimination mature by middle childhood? A study using time-frequency analysis of mismatch responses from 7 years to adulthood. *Developmental Science*, *14*(2), 402–416. <https://doi.org/10.1111/j.1467-7687.2010.00990.x>
- Boersma, P., & Weenink, D. (2011). *Praat: doing phonetics by computer* (Version 5.3.25).
- Carbajal, M. J., Peperkamp, S., & Tsuji, S. (2021). A meta-analysis of infants' word-form recognition. *Infancy*, *26*(3), 369–387. <https://doi.org/10.1111/inf.12391>
- Čeponien, R., Lepistö, T., Alku, P., Aro, H., & Näätänen, R. (2003). Event-related potential indices of auditory vowel processing in 3-year-old children. *Clinical Neurophysiology*, *114*(4), 652–661. [https://doi.org/10.1016/S1388-2457\(02\)00436-4](https://doi.org/10.1016/S1388-2457(02)00436-4)
- Chen, A., Peter, V., Wijnen, F., Schnack, H., & Burnham, D. (2018). Are lexical tones musical? Native language's influence on neural response to pitch in different domains. *Brain and Language*, *180–182*, 31–41. <https://doi.org/10.1016/j.bandl.2018.04.006>
- Chen, A., Wijnen, F., Koster, C., & Schnack, H. (2017). Individualized Early Prediction of Familial Risk of Dyslexia: A Study of Infant Vocabulary Development. *Frontiers in Psychology*, *8*, 156. <https://doi.org/10.3389/fpsyg.2017.00156>
- Cheng, Y.-Y., Wu, H.-C., Tzeng, Y.-L., Yang, M.-T., Zhao, L.-L., & Lee, C.-Y. (2015). Feature-specific transition from positive mismatch response to mismatch negativity in early infancy: Mismatch responses to vowels and initial consonants. *International Journal of Psychophysiology*, *96*(2), 84–94. <https://doi.org/proxy.library.uu.nl/10.1016/j.ijpsycho.2015.03.007>
- 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. In *Hearing Research* (Vol. 82, Issue 1, pp. 53–58). [https://doi.org/10.1016/0378-5955\(94\)00164-L](https://doi.org/10.1016/0378-5955(94)00164-L)
- Chládková, K., Escudero, P., & Lipski, S. C. (2015). When “AA” is long but “A” is not short: speakers who distinguish short and long vowels in production do not necessarily encode a short–long contrast in their phonological lexicon. *Frontiers in Psychology*, *6*, 438. <https://doi.org/10.3389/fpsyg.2015.00438>
- Choi, M., & Shukla, M. (2021). A New Proposal for Phoneme Acquisition: Computing Speaker-Specific Distribution. In *Brain Sciences* (Vol. 11, Issue 2). <https://doi.org/10.3390/brainsci11020177>
- Crinnion, A. M., Malmkog, B., & Toscano, J. C. (2020). A graph-theoretic approach to identifying acoustic cues for speech sound categorization. *Psychonomic Bulletin & Review*, *27*(6), 1104–1125. <https://doi.org/10.3758/s13423-020-01748-1>
- Cristia, A., & Seidl, A. (2014). The hyperarticulation hypothesis of infant-directed speech. *Journal of Child Language*, *41*(4), 913–934. <https://doi.org/10.1017/S0305000912000669>
- Dehaene-Lambertz, G., & Baillet, S. (1998). A phonological representation in the infant brain. *NeuroReport*, *9*(8), 1885–1888.
- Dehaene-Lambertz, G., & Pena, M. (2001). Electrophysiological evidence for automatic phonetic processing in neonates. *Neuroreport*, *12*(14), 3155–3158.
- 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>

- Dietrich, C., Swingle, D., & Werker, J. F. (2007). Native language governs interpretation of salient speech sound differences at 18 months. *Proceedings of the National Academy of Sciences*, *104*(41), 16027–16031. <https://doi.org/10.1073/pnas.0705270104>
- Englund, K. T., & Behne, D. M. (2005). Infant Directed Speech in Natural Interaction—Norwegian Vowel Quantity and Quality. *Journal of Psycholinguistic Research*, *34*(3), 259–280. <https://doi.org/10.1007/s10936-005-3640-7>
- Feldman, N. H., Griffiths, T. L., Goldwater, S., & Morgan, J. L. (2013). A role for the developing lexicon in phonetic category acquisition. *Psychological Review*, *120*(4), 751–778. <https://doi.org/10.1037/a0034245>
- Feldman, N. H., Myers, E. B., White, K. S., Griffiths, T. L., & Morgan, J. L. (2013). Word-level information influences phonetic learning in adults and infants. *Cognition*, *127*(3), 427–438. <https://doi.org/10.1016/j.cognition.2013.02.007>
- Fenson, L., Dale, P. S., Reznick, J. S., Bates, E., Thal, D. J., Pethick, S. J., Tomasello, M., Mervis, C. B., & Stiles, J. (1994). Variability in Early Communicative Development. *Monographs of the Society for Research in Child Development*, *59*(5), Variability in Early Communicative Development), i+iii-185.
- Fu, Z., & Monahan, P. J. (2021). Extracting Phonetic Features From Natural Classes: A Mismatch Negativity Study of Mandarin Chinese Retroflex Consonants. *Frontiers in Human Neuroscience*, *15*. <https://doi.org/10.3389/fnhum.2021.609898>
- Garcia-Sierra, A., Rivera-Gaxiola, M., Percaccio, C. R., Conboy, B. T., Romo, H., Klarman, L., Ortiz, S., & Kuhl, P. K. (2011). Bilingual language learning: An ERP study relating early brain responses to speech, language input, and later word production. *Journal of Phonetics*, *39*(4), 546–557. <https://doi.org/10.1016/j.wocn.2011.07.002>
- Gennari, G., Marti, S., Palu, M., Fló, A., & Dehaene-Lambertz, G. (2021). Orthogonal neural codes for speech in the infant brain. *Proceedings of the National Academy of Sciences*, *118*(31), e2020410118. <https://doi.org/10.1073/pnas.2020410118>
- Goldfield, B. A., & Reznick, J. S. (1990). Early lexical acquisition: rate, content, and the vocabulary spurt. *Journal of Child Language*, *17*(1), 171–183. <https://doi.org/10.1017/S0305000900013167>
- Goldinger, S. D., Pisoni, D. B., & Logan, J. S. (1991). On the nature of talker variability effects on recall of spoken word lists. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *17*(1), 152–162. <https://doi.org/10.1037/0278-7393.17.1.152>
- Harnsberger, J. D., Brown, W. S., Shrivastav, R., & Rothman, H. (2010). Noise and Tremor in the Perception of Vocal Aging in Males. In *Journal of Voice* (Vol. 24, Issue 5, pp. 523–530). <https://doi.org/10.1016/j.jvoice.2009.01.003>
- He, C., Hotson, L., & Trainor, L. J. (2009). Maturation of cortical mismatch responses to occasional pitch change in early infancy: Effects of presentation rate and magnitude of change. *Neuropsychologia*, *47*(1), 218–229. <https://doi.org/10.1016/j.neuropsychologia.2008.07.019>
- Hisagi, M., Garrido-Nag, K., Datta, H., & Shafer, V. L. (2015). ERP indices of vowel processing in Spanish–English bilinguals. *Bilingualism: Language and Cognition*, *18*(2), 271–289. <https://doi.org/10.1017/S1366728914000170>
- Hisagi, M., Shafer, V. L., Strange, W., & Sussman, E. S. (2010). Perception of a Japanese vowel length contrast by Japanese and American English listeners: Behavioral and electrophysiological measures. *Brain Research*, *1360*, 89–105. <https://doi.org/10.1016/j.brainres.2010.08.092>
- Johnson, K., & Sjerps, M. (2018). Speaker normalization in speech perception. *UC Berkeley Phonetics and Phonology Lab Annual Report*, 32–64.
- Kuhl, P. K. (1991). Human adults and human infants show a "perceptual magnet effect" for the prototypes of speech categories, monkeys do not. *Attention, Perception, & Psychophysics*, *50*(2), 93–107.
- Kuhl, P. K., Andruski, J. E., Chistovich, I. A., Chistovich, L. A., Kozhevnikova, E. V., Ryskina, V. L., Stolyarova, E. I., Sundberg, U., & Lacerda, F. (1997). Cross-Language Analysis of Phonetic Units in Language Addressed to Infants. *Science*, *277*(5326), 684–686. <https://doi.org/10.1126/science.277.5326.684>
- 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 of London. Series B, Biological Sciences*, *363* (1493), 979–1000. <https://doi.org/10.1098/rstb.2007.2154>
- Kushnerenko, E., van den Bergh, B., & Winkler, I. (2013). Separating acoustic deviance from novelty during the first year of life: a review of event-related potential evidence. *Frontiers in Psychology*, *4*. <https://doi.org/10.3389/fpsyg.2013.00595>

- Lee, C.-Y., Yen, H., Yeh, P., Lin, W.-H., Cheng, Y.-Y., Tzeng, Y.-L., & Wu, H.-C. (2012). Mismatch responses to lexical tone, initial consonant, and vowel in Mandarin-speaking preschoolers. *Neuropsychologia*, *50*(14), 3228–3239. <https://doi.org/10.1016/j.neuropsychologia.2012.08.025>
- Leppänen, P. H. T., Eklund, K. M., & Lyytinen, H. (1997). Event-related brain potentials to change in rapidly presented acoustic stimuli in newborns. *Developmental Neuropsychology*, *13*(2), 175–204. <https://doi.org/10.1080/87565649709540677>
- Mani, N., Coleman, J., & Plunkett, K. (2008). Phonological Specificity of Vowel Contrasts at 18-months. *Language and Speech*, *51*(1–2), 3–21. <https://doi.org/10.1177/00238309080510010201>
- Marklund, E., Schwarz, I.-C., & Lacerda, F. (2019). Amount of speech exposure predicts vowel perception in four- to eight-month-olds. *Developmental Cognitive Neuroscience*, *36*, 100622. <https://doi.org/10.1016/j.dcn.2019.100622>
- Martin, A., Schatz, T., Versteegh, M., Miyazawa, K., Mazuka, R., Dupoux, E., & Cristia, A. (2015). Mothers Speak Less Clearly to Infants Than to Adults: A Comprehensive Test of the Hyperarticulation Hypothesis. *Psychol Sci*, *26*(3), 341–347. <https://doi.org/10.1177/0956797614562453>
- Maurer, U., Bucher, K., Brem, S., & Brandeis, D. (2003). Development of the automatic mismatch response: from frontal positivity in kindergarten children to the mismatch negativity. In *Clinical Neurophysiology* (Vol. 114, Issue 5, pp. 808–817). [https://doi.org/proxy.library.uu.nl/10.1016/S1388-2457\(03\)00032-4](https://doi.org/proxy.library.uu.nl/10.1016/S1388-2457(03)00032-4)
- Maye, J., Werker, J. F., & Gerken, L. (2002). Infant sensitivity to distributional information can affect phonetic discrimination. *Cognition*, *82*(3), B101–B111. [https://doi.org/10.1016/S0010-0277\(01\)00157-3](https://doi.org/10.1016/S0010-0277(01)00157-3)
- McMurray, B., Danelz, A., Rigler, H., & Sedorff, M. (2018). Speech categorization develops slowly through adolescence. In *Developmental Psychology* (Vol. 54, Issue 8, pp. 1472–1491). American Psychological Association. <https://doi.org/10.1037/dev0000542>
- McMurray, B., Kovack-Lesh, K. A., Goodwin, D., & McEchron, W. (2013). Infant directed speech and the development of speech perception: Enhancing development or an unintended consequence? In *Cognition* (Vol. 129, Issue 2, pp. 362–378). <https://doi.org/10.1016/j.cognition.2013.07.015>
- Metsala, J. L., & Walley, A. C. (1998). *Spoken vocabulary growth and the segmental restructuring of lexical representations: Precursors to phonemic awareness and early reading ability* (J. L. M. L. C. Ehri, Ed.; pp. 89–120). Lawrence Erlbaum Associates Publishers.
- Morr, M. L., Shafer, V. L., Kreuzer, J. A., & Kurtzberg, D. (2002). Maturation of Mismatch Negativity in Typically Developing Infants and Preschool Children. *Ear and Hearing*, *23*(2), 118–136. http://journal.s.lww.com/ear-hearing/Fulltext/2002/04000/Maturation_of_Mismatch_Negativity_in_Typically.5.aspx
- Mulak, K. E., Bonn, C. D., Chládková, K., Aslin, R. N., & Escudero, P. (2017). Indexical and linguistic processing by 12-month-olds: Discrimination of speaker, accent and vowel differences. *PLOS ONE*, *12*(5), e0176762.
- Näätänen, R., Paavilainen, P., Rinne, T., & Alho, K. (2007). The mismatch negativity (MMN) in basic research of central auditory processing: A review. *Clinical Neurophysiology*, *118*(12), 2544–2590. <https://doi.org/10.1016/j.clinph.2007.04.026>
- Nazzi, T. (2005). Use of phonetic specificity during the acquisition of new words: Differences between consonants and vowels. *Cognition*, *98*(1), 13–30. <https://doi.org/10.1016/j.cognition.2004.10.005>
- Nazzi, T., & Bertoncini, J. (2003). Before and after the vocabulary spurt: two modes of word acquisition? *Developmental Science*, *6*(2), 136–142. <https://doi.org/10.1111/1467-7687.00263>
- Peltola, M. S., Kujala, T., Tuomainen, J., Ek, M., Aaltonen, O., & Näätänen, R. (2003). Native and foreign vowel discrimination as indexed by the mismatch negativity (MMN) response. *Neuroscience Letters*, *352* (1), 25–28. <https://doi.org/proxy.library.uu.nl/10.1016/j.neulet.2003.08.013>
- Poeppel, D., Guillemin, A., Thompson, J., Fritz, J., Bavelier, D., & Braun, A. R. (2004). Auditory lexical decision, categorical perception, and FM direction discrimination differentially engage left and right auditory cortex. *Neuropsychologia*, *42*(2), 183–200. <https://doi.org/10.1016/j.neuropsychologia.2003.07.010>
- Polka, L., & Bohn, O.-S. (2011). Natural Referent Vowel (NRV) framework: An emerging view of early phonetic development. *Journal of Phonetics*, *39*(4), 467–478. <https://doi.org/10.1016/j.wocn.2010.08.007>
- Rost, G. C., & McMurray, B. (2010). Finding the Signal by Adding Noise: The Role of Noncontrastive Phonetic Variability in Early Word Learning. *Infancy*, *15*(6), 608–635. <https://doi.org/10.1111/j.1532-7078.2010.00033.x>
- Ryalls, B. O., & Pisoni, D. B. (1997). The effect of talker variability on word recognition in preschool children. *Developmental Psychology*, *33*(3), 441–452. <https://doi.org/10.1037/0012-1649.33.3.441>

- Saloranta, A., Alku, P., & Peltola, M. S. (2020). Listen-and-repeat training improves perception of second language vowel duration: Evidence from mismatch negativity (MMN) and N1 responses and behavioral discrimination. *International Journal of Psychophysiology*, *147*, 72–82. <https://doi.org/10.1016/J.IJPSYCHO.2019.11.005>
- Schatz, T., Feldman, N. H., Goldwater, S., Cao, X.-N., & Dupoux, E. (2021). Early phonetic learning without phonetic categories: Insights from large-scale simulations on realistic input. *Proceedings of the National Academy of Sciences*, *118*(7), e2001844118. <https://doi.org/10.1073/pnas.2001844118>
- Shafer, V. L., Morr, M. L., Kreuzer, J. A., & Kurtzberg, D. (2000). Maturation of Mismatch Negativity in School-Age Children. *Ear and Hearing*, *21*(3), 242–251. http://journals.lww.com/ear-hearing/Fulltext/2000/06000/Maturation_of_Mismatch_Negativity_in_School_Age.8.aspx
- Shafer, V. L., Yu, Y. H., & Datta, H. (2010). Maturation of Speech Discrimination in 4- to 7-Yr-Old Children as Indexed by Event-Related Potential Mismatch Responses. *Ear and Hearing*, *31*(6), 735–745. https://journals.lww.com/ear-hearing/Fulltext/2010/12000/Maturation_of_Speech_Discrimination_in_4_to.2.aspx
- Shafer, V. L., Yu, Y. H., & Datta, H. (2011). The development of English vowel perception in monolingual and bilingual infants: Neurophysiological correlates. In *Journal of Phonetics* (Vol. 39, Issue 4, pp. 527–545). <https://doi.org/10.1016/j.wocn.2010.11.010>
- Shafer, V. L., Yu, Y. H., & Garrido-Nag, K. (2012). Neural mismatch indices of vowel discrimination in monolingually and bilingually exposed infants: Does attention matter? *Neuroscience Letters*, *526*(1), 10–14. <https://doi.org/10.1016/j.neulet.2012.07.064>
- Shestakova, A., Brattico, E., Huotilainen, M., Galunov, V., Soloviev, A., Sams, M., Ilmoniemi, R., & Näätänen, R. (2002). Abstract phoneme representations in the left temporal cortex: magnetic mismatch negativity study. *Neuro Report*, 1813–1816.
- Shestakova, A., Brattico, E., Soloviev, A., Klucharev, V., & Huotilainen, M. (2004). Orderly cortical representation of vowel categories presented by multiple exemplars. In *Cognitive Brain Research* (Vol. 21, Issue 3, pp. 342–350). <https://doi.org/10.1016/j.cogbrainres.2004.06.011>
- Singh, L. (2019). Does infant speech perception predict later vocabulary development in bilingual infants? *Journal of Phonetics*, *76*, 100914. <https://doi.org/10.1016/j.wocn.2019.100914>
- Sittiprapaporn, W., Tervaniemi, M., Chindaduangratn, C., & Kotchabhakdi, N. (2005). Preattentive discrimination of across-category and within-category change in consonant–vowel syllable. *Neuroreport*, *16*(13), 1513–1518.
- Swingle, D. (2009). Contributions of infant word learning to language development. *Philosophical Transactions of the Royal Society of London B: Biological Sciences*, *364*(1536), 3617–3632. <https://doi.org/10.1098/rstb.2009.0107>
- Swingle, D. (2019). Learning Phonology from Surface Distributions, Considering Dutch and English Vowel Duration. *Language Learning and Development*, *15*(3), 199–216. <https://doi.org/10.1080/15475441.2018.1562927>
- Swingle, D., & Alarcon, C. (2018). Lexical Learning May Contribute to Phonetic Learning in Infants: A Corpus Analysis of Maternal Spanish. *Cognitive Science*, *42*(5), 1618–1641. <https://doi.org/10.1111/cogs.12620>
- Trainor, L., McFadden, M., Hodgson, L., Darragh, L., Barlow, J., Matsos, L., & Sonnadora, R. (2003). Changes in auditory cortex and the development of mismatch negativity between 2 and 6 months of age. *International Journal of Psychophysiology*, *51*(1), 5–15. [https://doi.org/10.1016/S0167-8760\(03\)00148-X](https://doi.org/10.1016/S0167-8760(03)00148-X)
- Tsao, F.-M., Liu, H.-M., & Kuhl, P. K. (2004). Speech Perception in Infancy Predicts Language Development in the Second Year of Life: A Longitudinal Study. *Child Development*, *75*(4), 1067–1084. <https://doi.org/10.1111/j.1467-8624.2004.00726.x>
- Tuninetti, A., Chládková, K., Peter, V., Schiller, N. O., & Escudero, P. (2017). When speaker identity is unavoidable: Neural processing of speaker identity cues in natural speech. In *Brain and Language* (Vol. 174, pp. 42–49). <https://doi.org/10.1016/j.bandl.2017.07.001>
- van Leeuwen, T., Been, P., Kuijpers, C., Zwarts, F., Massen, B., & van der Leij, A. (2006). Mismatch response is absent in 2-month-old infants at risk for dyslexia. *Neuroreport*, *17*(16), 351–355.
- Walley, A., Metsala, J., & Garlock, V. (2003). Spoken vocabulary growth: Its role in the development of phoneme awareness and early reading ability. *Reading and Writing*, *16*(1–2), 5–20. <https://doi.org/10.1023/A:1021789804977>

- Wanrooij, K., Boersma, P., & Van Zuijlen, T. (2014). Fast phonetic learning occurs already in 2-to-3-month old infants: an ERP study. *Frontiers in Psychology*, 5, 77. <https://doi.org/10.3389/fpsyg.2014.00077>
- Werker, J. F., & Curtin, S. (2005). PRIMIR: A Developmental Framework of Infant Speech Processing. *Language Learning and Development*, 1(2), 197–234. <https://doi.org/10.1080/15475441.2005.9684216>
- Werker, J. F., Fennell, C. T., Corcoran, K. M., & Stager, C. L. (2002). Infants' Ability to Learn Phonetically Similar Words: Effects of Age and Vocabulary Size. *Infancy*, 3(1), 1–30. https://doi.org/10.1207/S15327078IN0301_1
- Xi, J., Zhang, L., Shu, H., Zhang, Y., & Li, P. (2010). Categorical perception of lexical tones in Chinese revealed by mismatch negativity. *Neuroscience*, 170(1), 223–231. <https://doi.org/10.1016/j.neuroscience.2010.06.077>
- Xiao, X.-Z., Wong, H. K., Wang, Y., Zhao, K., Zeng, G. Q., Yip, L.-Y., Wong, G. C.-S., & Tse, C.-Y. (2018). Detecting violation in abstract pitch patterns with mismatch negativity. *Psychophysiology*, 55(8), e13078. <https://doi.org/10.1111/psyp.13078>
- Yeung, H. H., & Werker, J. F. (2009). Learning words' sounds before learning how words sound: 9-Month-olds use distinct objects as cues to categorize speech information. *Cognition*, 113(2), 234–243. <https://doi.org/proxy.library.uu.nl/10.1016/j.cognition.2009.08.010>
- Yu, Y. H., Tessel, C., Han, H., Campanelli, L., Vidal, N., Gerometta, J., Garrido-Nag, K., Datta, H., & Shafer, V. L. (2019). Neural Indices of Vowel Discrimination in Monolingual and Bilingual Infants and Children. *Ear and Hearing*, 40(6), 1376–1390. <https://doi.org/10.1097/AUD.0000000000000726>
- Zink, I., & Lejaegere, M. (2002). *N-CDI Woordenlijst*. Acco.

Cite this article: Chen, A. (2024). Twenty-month-olds categorically discriminate similar sounding vowels regardless of vocabulary level, an event related potentials (ERP) study. *Journal of Child Language* 51, 434–453, <https://doi.org/10.1017/S0305000923000351>