Amount of speech exposure predicts vowel perception in four- to eight-month-olds

Ellen Marklund*, Iris-Corinna Schwarz, Francisco Lacerda

Stockholm University, Ellen Marklund, Department of Linguistics, SE-106 91, Stockholm, Sweden

ABSTRACT

During the first year of life, infants shift their focus in speech perception from acoustic to linguistic information. This perceptual reorganization is related to exposure, and a direct relation has previously been demonstrated between amount of daily language exposure and mismatch response (MMR) amplitude to a native consonant contrast at around one year of age. The present study investigates the same relation between amount of speech exposure and MMR amplitude to a native vowel contrast at four to eight months of age. Importantly, the present study uses spectrally rotated speech in an effort to take general neural maturation into account. The amplitude of the part of the MMR that is tied specifically to speech processing correlates with amount of daily speech exposure, as estimated using the LENA system.

1. Introduction

Early in development, from birth and up to around six months of age, infants are able to discriminate between most speech sounds used contrastively in the world's languages, regardless of their own native language (for reviews see Maurer and Werker, 2014; Werker and Curtin, 2005). At around eight months, infants no longer reliably discriminate between non-native vowels (Bosch and Sebastián-Gallés, 2003; Polka and Werker, 1994), and at around ten months, the same is valid for non-native consonants (Werker and Tees, 1984). During this time, they also become better able to discriminate native contrasts (Kuhl et al., 1992, 2006). These results are taken to indicate a "perceptual reorganization" of speech during the first year of life (Werker, 1995; Werker and Tees, 1984), or in other words, a transition from a language-general to a language-specific perception of speech sounds (Kuhl, 1994). That is, during the first year of life, infants' perception of speech sounds shifts from the acoustic characteristics of the sounds towards the linguistic information they convey.

Acoustic versus linguistic processing of sounds can be captured using the event-related potential (ERP) component mismatch negativity (MMN). The MMN is elicited in response to a violation of an expected rule or pattern of stimulation (Winkler, 2007). The simplest instance of this constitutes occasional presentations of an oddball, or deviant stimulus, among a series of identical standard stimuli. Detection of the deviant stimulus results in a negative deflection in the ERP difference wave, that is, in the averaged response to the deviant stimulus minus the averaged response to the standard stimulus. That negative deflection, occurring around 150–250 ms after change onset, is the MMN (e.g., Näätänen et al., 2007). The amplitude of the MMN component reflects both degree of acoustic difference between standard and deviant stimulus (e.g., Tiitinen et al., 1994) and whether the two stimuli belong to the same or different perceptual categories (Winkler et al., 1999; Sharma and Dorman, 1999). For example, the same two speech stimuli induce a larger MMN in a group of participants to whom the difference between the two sounds is contrastive, compared to in a group in which they belong to the same speech sound category (Winkler et al., 1999).

In infants, the mismatch response (MMR) is not as well studied as in adults. Different experimental conditions result in different MMR morphologies, with varying latencies, amplitudes, and polarity (e.g., Cheour et al., 2001). However, the MMR has been demonstrated repeatedly to reflect the transition from acoustic to linguistic processing of speech sounds (e.g., Cheour et al., 1998; Rivera-Gaxiola et al., 2005). For example, the MMR to two vowels reflects only the acoustic difference between two vowels in a group of six-month-olds, whereas at twelve months of age, the same stimuli reflect enhanced MMRs to good

exemplars of native vowels (Cheour et al., 1998).

The transition from acoustic towards linguistic processing of speech has been tied to language exposure. Numerous studies have established that the sensitivity to contrasts in non-native languages is attenuated (e.g., Werker and Tees, 1984; Polka and Werker, 1994), whereas sensitivity to native contrasts is enhanced (Kuhl et al., 1992, 2006). Infants with two (or more) native languages typically retain the ability to discriminate contrasts that are relevant in either language (Burns et al., 2007). The ability to discriminate a non-native contrast can be retained beyond the age at which it typically attenuates, with relatively brief live exposure to a non-native language in which the contrast is relevant (Kuhl et al., 2003; Conboy and Kuhl, 2011). Furthermore, infants' perception of speech sounds can be influenced by only minutes of exposure to different types of frequency distributions conveying speech sound category information (Maye et al., 2002; Yoshida et al., 2010).

The amount of language exposure in infants' everyday life has also been tied to more mature MMRs both on a group level (Garcia-Sierra et al., 2011) and on an individual level (Garcia-Sierra et al., 2016). In the latter study, the MMRs of 11- to 14-month-old monolingual and bilingual children were related to the amount of daily language exposure. A large amount of daily language exposure was tied to a late negative MMR, whereas a somewhat smaller amount of language exposure was tied to an early positive MMR. The group with the smallest amount of exposure displayed no MMR in either the early or the late time-window (Garcia-Sierra et al., 2016). A negative MMR can be considered a more mature mismatch response than a positive MMR, since the polarity of the adult MMN is negative (Näätänen et al., 2007). That this response has longer latency in infant and child ERPs than it does in adult ERPs is in line with development of other ERP components (Shibasaki and Miyazaki, 1992).

The present study expands the findings of Garcia-Sierra and colleagues (Garcia-Sierra et al., 2011; 2016), by investigating the relation between daily language exposure and the MMR to native speech sound contrasts at younger ages than previously studied. Further, a non-speech control condition was included since the morphology of the MMR undergoes drastic general changes across development (Cheour et al., 2001). Spectrally rotated speech was used as non-speech control because it shares many acoustic characteristics with speech and is of comparable acoustic complexity (Marklund et al., 2018), while it does not sound anything like speech. Language-unrelated developmental differences in MMR realization can be accounted for by including this condition, since infants are not exposed to spectrally rotated speech in their everyday life.

2. Method

2.1. Participants

A total of 42 infants (20 girls, 22 boys) participated in the study. From this sample, fifteen infants were excluded from the final analysis, either due to technical problems during the EEG-recording session (n = 3), or because less than 30 trials remained in one or both deviant conditions after data pre-processing (n = 8), or because they were outliers in terms of MMR amplitude (n = 4). Subjects were classified as outliers if the mean voltage in one or more MMR measures differed more than 2 SD from the group mean of that measure (details on the four MMR measures can be found in section *EEG preprocessing* below). This resulted in 12 participants in the younger age group (mean age 4.4 months, SD = 0.5) and 15 participants in the older age group (mean age 8.1 months, SD = 0.7). After preliminary analyses, the two age groups were combined, resulting in a total number of 27 participants (14 girls, 13 boys) in the final analysis, with a mean age of 6.4 months (SD = 2.0). Participant age is reported based on the date of the EEG experiment. Audio recordings in the home took place on average four days after the experiment date (SD = 3; range = 1–20).

All included infants (n = 27) were learning Swedish as a first

language, with 18 being monolingual Swedish-learning infants. The remaining nine families were bilingual (n = 6) or trilingual (n = 3), with Swedish as one of the languages spoken with the infant. In one bilingual and one trilingual family both parents reported that they only speak Swedish to their child, not their own first language. Those subjects are nevertheless included in the multilingual group for the purposes of this study, since they are most likely at least passively exposed to multiple languages in their daily life. The vowels used in the experiment (/e/ and /i/) were part of the phoneme inventory in at least two of the multilingual children's native languages.

The socioeconomic status (SES), estimated by the mothers' level of education, was high and homogenous within the included participants; 89% (24/27) of the mothers had university education or equivalent. Two of the remaining mothers had higher vocational education, and the third did not provide information about educational level. In all families, one or both parents were on parental leave, totaling full time (i.e., none of the participants were enrolled in pre-school[1]).

Parents of participants received a small children's book and a photo of their child with the EEG-net as thanks for their participation in the study. The study has been approved by the Ethical Review Board at Karolinska Institutet (2015/63-31).

2.2. Stimuli

The stimuli in the MMR experiment were two vowels spoken by a female native speaker of Swedish, and the rotated counterparts of the same vowels. The vowels, /e/ and /i/, were recorded in /Vt/ and /Vk/ syllable contexts. Recording took place in an anechoic chamber equipped with a Nexus Conditioning Amplifier (Brüel & Kjær Sound & Vibration Measurement A/S, Nærum, Denmark), Brüel & Kjær microphones (model 4189), and Adobe Audition 1.5 (Adobe Systems Inc., San Jose, CA, USA). Twelve exemplars of each syllable were recorded. Listening to the recordings, two exemplars of each vowel were selected as possible stimuli based on auditory similarity. Their consonant context was subsequently removed, and the acoustic properties of the remaining vowels were mapped using Wavesurfer 1.8.5 (Sjölander and Beskow, 2000). One /e/ and one /i/ were chosen as stimuli based on the fact that they were approximately matched for fundamental frequency (f_o), intensity, and duration (see Table 1).

The two vowels were spectrally rotated to create the rotated speech stimuli, a procedure that entails multiplying the speech signal with a carrier wave so that the negative part of the spectrum is shifted to above zero, then using a low-pass filter to remove the positive part of the spectrum. Carrier wave frequency and filter cut-off frequency determine the frequency span included in the spectral rotation (Blesser, 1972; Marklund et al., 2018). The rotation script was written in Mathematica 9 (Wolfram Research Inc., Champaign, Illinois, USA), and the frequency span included in the rotation was 0-5000 Hz.

Four adult participants took part in an active oddball perceptual test with the selected stimuli in order to make sure both the vowels and their rotated counterparts were discriminable. For the natural vowels, accuracy was 100.0% and the false positives rate was 0.0%. For the rotated vowels, accuracy was 91.7% and the false positives rate was 0.4%.

2.3. Experiment design and procedure

The experiment was of a classic passive oddball design with 16 blocks, each containing 120 trials, with 10% deviants. Each block contained either speech sounds or rotated speech sounds. For half of the

---

[1] It is possible that participants attended" open pre-school", an activity similar to pre-school that parents and infants/children can attend together for a few hours per day. Parents were not asked whether they regularly attended "open pre-school" or other organized activities with their child.

Table 1
Acoustic information about the speech stimuli used in the MMR experiment.

Vowel	Mean f_o	f0-contour	Mean intensity	Intensity contour	Duration
/e/	182 Hz		40 dB		350 ms
/i/	175 Hz		45 dB		350 ms

Table 2
Number of participants in the counterbalancing groups. Numbers in parentheses indicate number of participants before exclusion criteria were applied.

	Standard /e/	Standard /i/
First block speech sounds	n = 6 (11)	n = 6 (11)
First block rotated sounds	n = 7 (10)	n = 8 (10)

participants, the first block presented speech sounds and the second block presented rotated sounds, alternating the two block types throughout the experiment in order to get as even a distribution between speech and rotated speech trials as possible. The other half of the participants heard rotated speech sounds in the first block and speech sounds in the second block, and so on. For half of the subjects, the /e/ was presented as standard in the speech sound blocks, and /i/ as the deviant. For the same participants, the rotated /e/ served as standard and the rotated /i/ served as deviant in the rotated speech sound blocks. For the other half of the participants, the /i/ and the rotated /i/ were standards and the /e/ and the rotated /e/ were deviants in their respective blocks. The reason for counterbalancing standard and deviant vowels is that the MMN amplitude reflects asymmetries in vowel discrimination (Molnar et al., 2014). After excluding participants, the distribution of subjects was still fairly balanced (see Table 2).

There were at least twelve standard stimuli at the beginning of each block, and at least three standards preceding each deviant. Stimuli duration was 350 ms and the inter-stimulus interval was 650 ms, resulting in a trial duration of 1000 ms. Between the blocks were short pauses in order to increase the possibility of keeping infant participants engaged in the experiment. Infants were seated on their parent's lap and silently entertained by a research assistant. Both the parent and the research assistant listened to masking music in headphones. Each block lasted for about two minutes and pauses between blocks were 30 s long, resulting in a total experiment duration of approximately 45 min. However, most infants became fussy at some point in which case the experiment was terminated early.

2.4. EEG recording

The EEG-system consists of a NetAmps 300 amplifier (20 kHz sampling rate), Hydro-Cel 124-electrode high-impedance nets for infants, and EGI NetStation 4.4.2 software for recording (Electrical Geodesics Inc. Eugene, Oregon, USA). Stimuli were presented using E-Prime 2.0 (Psychology Software Tools, Sharpsburg, Pennsylvania, USA). During recording, reference was Cz, and the EEG signal was low-pass filtered online at 4 kHz. The signal was also down-sampled online during recording to a sampling rate of 250 Hz.

2.5. EEG preprocessing

The EEG-data was exported from the NetStation format using the NetStation Exporter application (Electrical Geodesics Inc.). The files were imported into the EEGLAB toolbox (Delorme and Makeig, 2004) for MATLAB R2014b (MathWorks Inc., Natick, Massachusetts, USA). The data was band-pass filtered at 1–20 Hz, re-referenced to the average of the two mastoid electrodes (electrode 57 and 100 in the EGI 124-channel nets), and epoched to -100 ms to 900 ms around stimulus onset. Baseline correction was applied to all epochs. The data was then preprocessed using the ERP PCA Toolkit 2.54 (Dien, 2010) for MATLAB. Bad channels were identified (max-min > 300 μV) and replaced when possible, movement artifacts were identified and removed using principal-component analysis (PCA; components removed if exceeding 200 μV), and trials containing more than 25% channels in which the span exceeded 300 μV were excluded. Infants with recordings in which less than 30 trials remained in any of the deviant conditions were excluded from further analysis. Subject difference waveforms were calculated by subtracting the subject average standard waveform (all standards included) from the subject average deviant waveform. The electrode site selected for analysis was Fz (in the present study represented by the average of electrodes 11, 16, 19 and 4 in the EGI 124-channel nets), since the MMN response in adults is typically strongest there (Näätänen et al., 2007). It could be argued that left frontal electrodes, for example F3, should be used since the MMR is typically strongest there in infant studies (e.g., Garcia-Sierra et al., 2011; 2016). However, since it is the adult MMN component that has been linked to discrimination (Tiitinen et al., 1994) and the relationship between the MMN and the infant MMR is not straightforward (Kushnerenko et al., 2013), the Fz site was selected for analysis in the present study on theoretical bases. The average amplitude of the difference waveform was calculated for two time-windows: 150–350 ms and 350–550 ms after stimulus onset (M150 and M350 respectively, M for mismatch response, 150 and 350 for the onsets of the analysis time-windows). The time-windows were chosen in line with a similar recent study (Garcia-Sierra et al., 2016).

2.6. Audio recordings and segmentation

Audio recordings were performed in the home, using recording devices from the Language Environment Analysis system (LENA; LENA Research Foundation, Boulder, CO, USA). At the time of the laboratory visit for the MMN experiment, parents were given two recording devices, one for each day of recording. They were also given two vests to place the devices in during recordings, and instructions on how to use them. They were asked to dress the child in the vest as he or she woke up on the day of the recording, put the device on and let it run for the rest of the day. During bath-time and nap-time the vest was removed but kept close by with the recording still running. The parents were told to inform other people entering their home on the day of the recording that the recording was going on. Parents also had the option to select parts of the recordings to be removed after the initial automated audio analysis. Each participating family was asked to perform two daylong recordings, one during a weekday and one during a weekend day. In some cases, the families were unable to perform recordings according to this pattern, in which case they performed two recordings on weekdays (n = 8) or two recordings on weekends (n = 1). In one case, the family contributed with a single weekday recording. The recordings were made on days that the infant and a parent primarily spent at home.

2.7. Automatic audio segmentation

The LENA-Pro 3.4.0 software (LENA Research Foundation, Boulder, CO, US) performs an automatic segmentation of the recording upon downloading it from the recording devices. Based on acoustic characteristics of the signal, the software identifies sections of the recordings with target child vocalizations, other child vocalizations, female speech, male speech, silence, TV/electronic sounds, and noise, including overlapping speech and other sounds that the system does not recognize (Xu et al., 2008). The amount of speech exposure was calculated for each subject by taking the average amount of exposure (in seconds) from the two recording days. In the case where only one recording was available, the estimates from this recording were used.

Table 3
The speech measures for the two age groups. Amount of exposure is reported in minutes.

Age group	N	Mean (SD)	Max	Min
4-month-olds	12	82.8 (32.7)	145.4	46.3
8-month-olds	15	74.8 (38.7)	142.1	23.6

3. Results

3.1. Speech measures

Although the amount of speech exposure is exported from the LENA system in seconds, and all statistical analyses were performed using this unit, results will be reported and visualized in minutes.

A paired-samples two-tailed *t*-test revealed no difference in total amount of speech exposure between weekdays and weekend days (t (27) = .053, p = .958). Data from all participants who had contributed with one weekend and one weekday recording (n = 28), regardless of whether their MMR recording was included or not, were included in this preliminary analysis. In the main analysis, the mean of the two contributed recordings was used as speech exposure measure, regardless of whether the recording had been done on a weekday or on a weekend day. In the case where only one recording was available, data from that recording was used.

The mean speech exposure amount was 78.4 min per day (SD = 35.7). There was no significant difference in speech exposure between the two age groups (t(25) = .567, p = .576; Table 3).

3.2. Mismatch responses

The grand average standard, deviant and difference waveforms at the Fz site can be found in Fig. 1. Fig. 2 shows the topography of the speech and rotated speech conditions in both analysis time-windows. To measure whether significant MMRs were found, one sample t-tests on the MMR amplitude (mean of all samples within the time-window) were performed (H$_0$ = 0), one per condition and analysis time-window. The amplitude of the difference waves differed significantly from zero in the earlier analysis window for both conditions, but only in the rotated condition in the later time-window (see Table 4). There was no difference in number of standard or deviant trials between conditions (deviants: t(26) = .959, p = .346, standards: t(26) = 1.213, p = .236).

Preliminary analyses were performed to investigate the effect of age, infant sex, standard vowel, block order and number of native languages on the MMR amplitudes. The preliminary analyses consisted of a series of repeated measures ANOVAs, all with MMR amplitudes in the two time-windows as dependent variables and speech condition (speech vs. rotated speech) as within-subject variable. In each of the five separate ANOVAs, one of the potentially confounding variables was entered as a between-subject variable. Ideally, all variables should of course have been entered in the same model, but due to the low sample size, such a complete model would have very low statistical power. Therefore, in order not to risk missing any effects of the potentially confounding variables, separate tests were run for each variable. For the same reason, the alpha level was not adjusted for multiple comparisons.

The preliminary analyses revealed a marginally significant interaction (p = .076) between time-window (M150 and M350) and age group (4-month-olds and 8-month-olds), and a significant interaction (p = .036) between condition (speech and rotated speech) and sex (girls and boys). No significant interactions were found for standard vowel, block order or number of native languages. Details about the initial preliminary analysis can be found in Table 5.

The first finding of the preliminary analyses, the interaction between time-window and age group, reveals that there is a difference between the realization of the MMR in the two age groups that is *unrelated to condition* (see Fig. 3, left panel). In order to isolate the effect of speech exposure on speech sound category development (quantified as MMR amplitude), this age-related difference in MMR realization in general must first be accounted for. Based on the reasoning that an MMR to a change between two speech sounds is comprised of two separate comparison processes (Cheour et al., 2001), and the fact that the MMR to two rotated speech sounds has been shown to reasonably approximate the acoustic part of the MMR response (Marklund et al., 2018), the part of the MMR specifically tied to language processing was estimated by subtracting the MMR to the rotated speech from the MMR to the vowels (separately for each time-window).

In other words, preliminary analysis revealed that there are age-related differences in the MMRs that are found across both the speech and the rotated speech condition. These differences are presumably unrelated to the processing of speech specifically, and instead related to the development of more general acoustic processing. Therefore, an approximation of the part of the MMR that is tied to acoustic processing is subtracted from the total MMR, isolating the part tied to language processing.

Follow-up analyses were performed to test for effects of age group on this isolated language-related part of the MMR. A repeated measures ANOVA with a corresponding design as in the original preliminary analyses was used, except that the dependent variable was the difference measure for each time window (henceforth called D150 and D350;
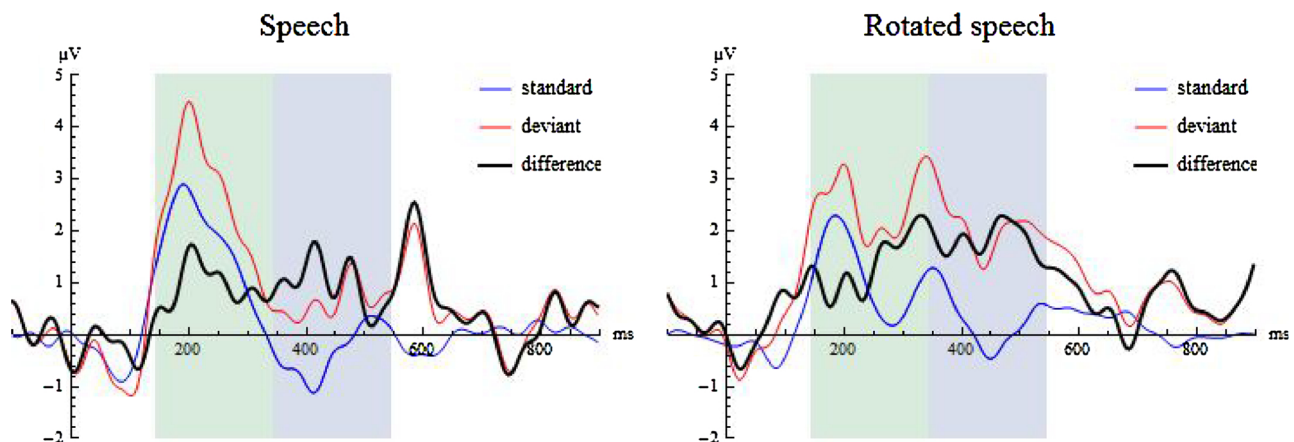

Fig. 1. The grand average standard, deviant and difference waveforms from the Fz site. The shaded green area shows the early analysis time-window (150–350 ms), and the shaded blue area shows the late analysis time-window (350–550 ms). Negativity is plotted down (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article).
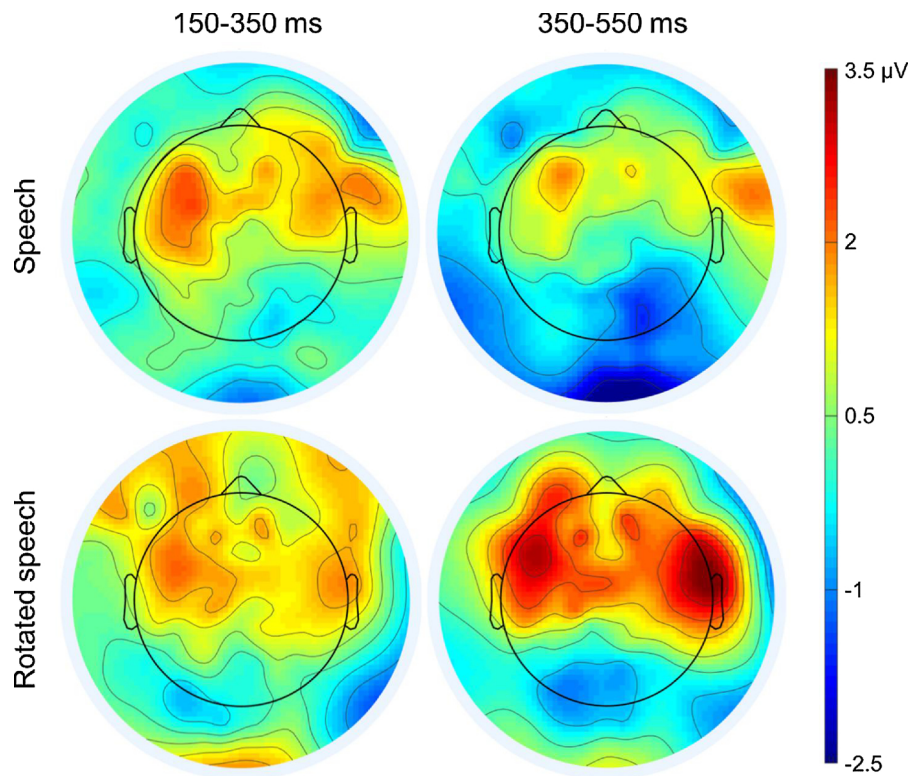
Fig. 2. The topography of the MMRs for the two conditions (rows) in the two analysis time-windows (columns).

Table 4
Results of t-tests of the MMR for the two conditions in the two time-windows at Fz (average of electrodes 11, 16, 19 and 4). Significant results are marked with an asterisk ($p < .05$).

Condition	Time-window	t	df	p
Speech	150-350 ms	2.127	26	.043*
	350-550 ms	0.921	26	.365
Rotated speech	150-350 ms	2.540	26	.017*
	350-550 ms	3.014	26	.006*

Table 5
Results of preliminary analyses testing for effects of potentially confounding factors. Significant results are marked with an asterisk ($p < .05$). Asterisk within parentheses denotes non-significant result but low enough p-value to warrant further investigation anyway ($p < .1$).

Between-subject factor	Interaction	F	df	p
Age group	Condition*Age group	.459	1, 25	.504
	Time-window*Age group	3.426	1, 25	.076(*)
Infant sex	Condition*Infant sex	4.919	1, 25	.036*
	Time-window*Infant sex	.150	1, 25	.702
Standard vowel	Condition*Standard vowel	.041	1, 25	.840
	Time-window*Standard vowel	.150	1, 25	.702
Block order	Condition*Block order	.711	1, 25	.407
	Time-window*Block order	.693	1, 25	.413
Lingualism	Condition*Lingualism	1.614	1, 25	.216
	Time-window*Lingualism	.378	1, 25	.544

D for amplitude difference) instead of the MMR amplitudes, and condition was not included as a within-subject variable. No interaction with age group was found using this measure (see also Fig. 3, right panel).

The second finding of the preliminary analyses was an interaction between condition and infant sex (Table 5). Probing this interaction using independent-samples t-tests showed that for speech, both the M150 amplitude ($p = .035$) and the M350 amplitude ($p = .024$)

differed between boys and girls, whereas for rotated speech neither MMR amplitude differed between the sexes (Table 6). Using the measures of the isolated language-related response (D150 and D350), an overall amplitude difference was found between boys and girls ($p = .036$). Follow-up t-tests revealed that the overall difference was driven by a difference in D150 ($p = .044$), and that there was no difference in D350 between boys and girls.

In the main analysis, only the difference measures were used in order to investigate the impact of speech exposure specifically on the part of the MMR that is related to language. Further, since no effect of age was evident in this measure, the two age groups were combined. Linear regressions showed that total amount of speech exposure predicted D350 ($F(1, 25) = 8.416$, $\beta = .502$, $R^2 = .252$, $p = .008$), but not D150 ($F(1, 25) = 2.003$, $\beta = .272$, $R^2 = .074$, $p = .169$), see Fig. 4. Separate analyses for boys and girls were not possible because of the small sample size. On the level of individual participants, a high positive D350 value is taken to indicate high specialization for language processing (higher MMR amplitude for speech than for rotated speech) and a low or negative D350 value indicates low language processing specialization (equal MMR amplitude or higher amplitude for rotated speech than for speech)[2].

All statistical tests were performed in SPSS 21 (International Business Machines Corp., Armonk, New York, USA).

---

[2] Previous studies have demonstrated a relation between a more negative MMR in this late time-window and both more exposure (Garcia-Sierra et al., 2016) and later language outcomes (Kuhl et al., 2008). However, this was found in older infants, where the MMR presumably was negative on a group level. In the present study it is assumed that a larger perceived difference should result in a larger MMR amplitude, regardless of polarity, based on the notion that the MMR at least to some extent is a precursor to – and shares characteristics of – the adult MMN (Kushnerenko et al., 2013). In the present sample, all measurable MMRs were positive, and a higher positive amplitude in the speech condition is thus taken as more specialized processing. See Discussion for problematization of the developmental transition from positive to negative MMRs.
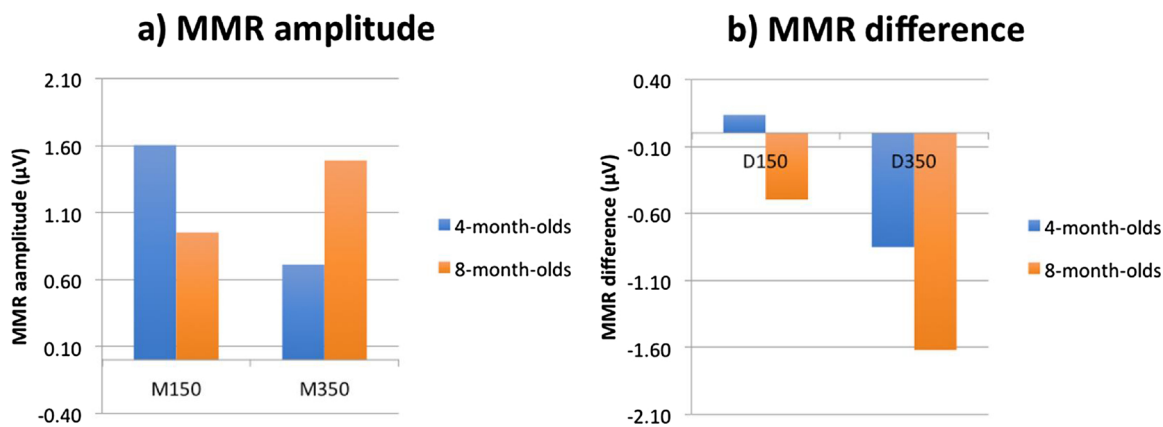
a) MMR amplitude

b) MMR difference

Fig. 3. a) The mean M150 and M350 amplitudes (across conditions) for 4-month-olds and 8-month-olds illustrate the marginally significant interaction between analysis time-window and age group. b) There is no significant interaction between analysis time-window and age group for the corrected MMRs (D150 and D350), in which the MMRs to rotated speech have been subtracted from the MMRs to vowels.

Table 6
The MMR amplitudes and the difference measures in both time windows, for all participants (n = 27), and divided into boys (n = 13) and girls (n = 14). MMR amplitudes and the difference measure (SD in parentheses) are given in μV.

Condition/measure	Infant sex	150-350 ms	350-550 ms
MMR speech	girls	2.20 (2.45)	1.67 (3.02)
	boys	−0.01 (2.71)	−0.75 (2.08)
	all	1.13 (2.77)	0.50 (2.84)
MMR rotated speech	girls	1.35 (2.82)	1.99 (2.85)
	boys	1.35 (2.81)	1.56 (3.40)
	all	1.35 (2.76)	1.78 (3.07)
Difference measure	girls	0.85 (2.71)	−0.32 (3.32)
	boys	−1.36 (2.70)	−2.31 (2.85)
	all	−0.22 (2.88)	−1.27 (3.21)

4. Discussion

A positive MMR was found in the earlier time-window (150–350 ms) for both the speech condition and the rotated speech condition. This is in line with previous findings of a positive MMR in this time-window both for native speech sound contrasts (Garcia-Sierra et al., 2016) and for non-speech sounds such as sine tones (Morr et al., 2002). The early MMR to speech contrasts is viewed as a change-detection response relying on attention to stimuli (Shafer et al., 2011; Garcia-Sierra et al., 2016). It has also been suggested to reflect intrinsic factors of the infant, such as general discriminatory ability and level of neuronal maturation, rather than extrinsic factors such as familiarity with (or automatized processing of) the specific sounds (Shafer et al., 2011). In the present study, no amplitude difference was found between the speech and the rotated speech condition, which is in line with this reasoning. Firstly, the general attention to speech stimuli and non-speech stimuli, quantified as the number of accepted trials in each condition, did not differ between conditions. Second, the rotated vowel stimuli were chosen based on the fact that an acoustic difference between two rotated sounds is a good approximation of the acoustic difference between the two original, un-rotated sounds (Marklund et al., 2018). It is thus not unexpected that there is no difference in early MMR amplitude (which reflects general discriminatory ability) between the two conditions.

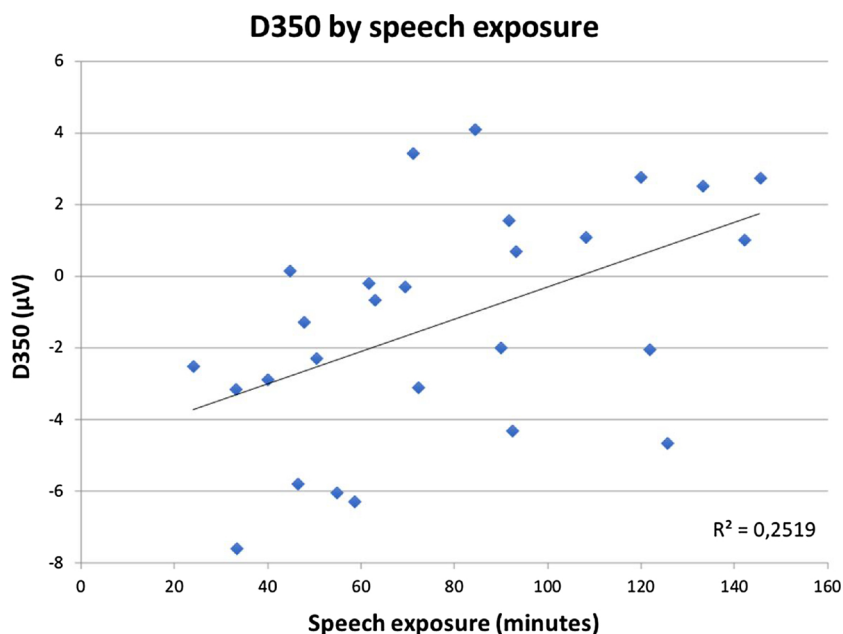In the later time-window (350–550 ms), a positive MMR was found

D350 by speech exposure

Fig. 4. The D350 displayed as a function of total amount of speech exposure. More positive D350 values, that is greater difference between the speech condition and the rotated speech condition, indicate higher degree of language specialization.

only for the rotated speech condition, but not for the speech sounds. The findings of Garcia-Sierra et al. (2016) show that a high amount of language exposure is related to a negative MMR in this time-window. In children with a moderate or low amount of language exposure, no MMR to native contrasts was found in this time-window. Assuming that the MMR to the rotated speech reasonably approximates a hypothetical MMR to speech sounds if the infants had never before been exposed to speech, it is possible that the lack of a measurable late MMR in the speech condition is in fact a sign of the impact of exposure on the MMR amplitude. If the MMR shifts gradually from positive with little or no exposure (approximated by the rotated speech condition, and the non-native Mandarin Chinese contrast in Garcia-Sierra et al., 2016) to negative with a lot of exposure (the high-exposure monolingual group in Garcia-Sierra et al., 2016), then it must at some point at this path pass zero. This is problematic, as whenever the amplitude of an MMR is not different from zero, it can also signify that the change was not detected.

The transition from a positive MMR at four to eight months of age (present study) to a negative MMR at around one year of age (Garcia-Sierra et al., 2016, high-exposure monolingual group) is most likely also at least in part a result of brain maturation in general, since the adult MMN in response to never before heard non-speech sounds is negative (Marklund et al., 2018). Including non-speech control conditions when studying speech sound perception developmentally can help shed light on how brain maturation and language exposure interact in shaping the infant MMR, which in turn is vital in order to interpret the MMRs in a meaningful way.

The age difference in how the MMR is realized in the early versus the late time-windows is not statistically significant in the present study, although still important. Since there was no interaction with condition, it illustrates the necessity of a non-speech condition of comparable complexity to speech when studying speech sound perception developmentally. Subtracting the MMRs to the rotated condition from the MMRs of the speech conditions isolates an approximation of the part of the MMRs that is specifically tied to changes in speech sound perception by removing the part of the MMR that is elicited by changes in any type of auditory stimuli. Subtracting the general acoustic part of the MMRs also removed the condition-independent difference between age groups, suggesting that speech sound processing development is not necessarily heavily influenced by maturational factors between four and eight months of age. Instead, this corrected measure was found to be directly related to daily language exposure. Importantly, this relation was found only for the later analysis time-window (350–550 ms). This is in line with previous findings showing an effect of amount of exposure on the MMR amplitude in this late time-window (Garcia-Sierra et al., 2016).

A difference was found in MMR amplitude between boys and girls. This is in line with previous findings, in which girls' brain responses have been demonstrated to mature more rapidly during infancy (e.g., Shucard and Shucard, 1990). The difference was found both in the early and the late MMRs in the speech condition, but not in the rotated speech condition. Interestingly, after isolating the part of the MMRs specific to speech sound perception (by subtracting the MMRs to rotated speech from the MMRs to speech), the difference between boys and girls remained but only in the early time-window. Since MMR amplitude in the early time-window has been tied to attention, general discriminatory ability and/or maturation (Garcia-Sierra et al., 2016; Shafer et al., 2011), this is in line both with findings showing faster maturation of speech related areas for girls than for boys (Shucard and Shucard, 1990), and with findings suggesting that female infants attend to speech stimuli to a greater extent than male infants do (Shafer et al., 2012). It is worth noting that if the higher positive MMR amplitude for girls is an indication of attention, it is realized differently than in the study by Shafer et al. (2012), in which it was realized as a more negative MMR in later time-windows.

In conclusion, the present study shows that four-to eight-month-old infants' processing of native speech sounds is directly related to the amount of speech the infants are exposed to, more so than it is related to age-dependent maturational factors. More research is necessary to disentangle the relative impact of maturational factors, language exposure and development, as well as attentional factors on the infant MMR.

Conflict of Interest

None.

References

Blesser, B., 1972. Speech perception under conditions of spectral transformation: I. Phonetic characteristics. Journal of Speech. J. Speech Lang. Hear. Res. 15 (1), 5–41. https://doi.org/10.1044/jshr.1501.05.

Bosch, L., Sebastián-Gallés, N., 2003. Simultaneous bilingualism and the perception of a language-specific vowel contrast in the first year of life. Lang. Speech 46 (2-3), 217–243. https://doi.org/10.1177/00238309030460020801.

Burns, T.C., Yoshida, K.A., Hill, K., Werker, J.F., 2007. The development of phonetic representation in bilingual and monolingual infants. Appl. Psycholinguist. 28 (3), 455–474. https://doi.org/10.1017/S0142716407070257.

Cheour, M., Ceponiene, R., Lehtokoski, A., Luuk, A., Allik, J., Ahlo, K., Näätänen, R., 1998. Development of language-specific phoneme representations in the infant brain. Nat. Neurosci. 1 (5), 351–353.

Cheour, M., Korpilahti, P., Martynova, O., Lang, A.H., 2001. Mismatch negativity and late discriminative negativity in investigating speech perception and learning in children and infants. Audiol. Neurotol. 6 (1), 2–11.

Conboy, B.T., Kuhl, P.K., 2011. Impact of second-language experience in infancy: brain measures of first-and second-language speech perception. Dev. Sci. 14 (2), 242–248. https://doi.org/10.1111/j.1467-7687.2010.00973.x.

Delorme, A., Makeig, S., 2004. EEGLAB: an open source toolbox for analysis of single-trial EEG dynamics including independent component analysis. J. Neurosci. Methods 134 (1), 9–21. https://doi.org/10.1016/j.jneumeth.2003.10.009.

Dien, J., 2010. The ERP PCA Toolkit: an open source program for advanced statistical analysis of event-related potential data. J. Neurosci. Methods 187 (1), 138–145. https://doi.org/10.1016/j.jneumeth.2009.12.009.

Garcia-Sierra, A., Rivera-Gaxiola, M., Percaccio, C.R., Conboy, B.T., Romo, H., Klarman, L., et al., 2011. Bilingual language learning: An ERP study relating early brain responses to speech, language input, and later word production. J. Phon. 39 (4), 546–557. https://doi.org/10.1016/j.wocn.2011.07.002.

Garcia-Sierra, A., Ramírez-Esparza, N., Kuhl, P.K., 2016. Relationships between quantity of language input and brain responses in bilingual and monolingual infants. Int. J. Psychophysiol. 110, 1–17. https://doi.org/10.1016/j.ijpsycho.2016.10.004.

Kuhl, P.K., 1994. Learning and representation in speech and language. Curr. Opin. Neurobiol. 4 (6), 812–822. https://doi.org/10.1016/0959-4388(94)90128-7.

Kuhl, P.K., Williams, K.A., Lacerda, F., Stevens, K.N., Lindblom, B., 1992. Linguistic experience alters phonetic perception in infants by 6 months of age. Science 255 (5044), 606–608. https://doi.org/10.1126/science.1736364.

Kuhl, P.K., Tsao, F.-M., Liu, H., 2003. Foreign-language experience in infancy: effects of short term exposure and social interaction on phonetic learning. Proc. Natl. Acad. Sci. 100 (15), 9096–9101. https://doi.org/10.1073/pnas.1532872100.

Kuhl, P.K., Stevens, E., Hayashi, A., Deguchi, T., Kiritani, S., Iverson, P., 2006. Infants show a facilitation effect for native language phonetic perception between 6 and 12 months. Dev. Sci. 9 (2), F13–F21. https://doi.org/10.1111/j.1467-7687.2006.00468.x.

Kuhl, P.K., Conboy, B.T., Coffey-Corina, S., Padden, D.M., Rivera-Gaxiola, M., Nelson, T., 2008. Phonetic learning as a pathway to language: new data and native language magnet theory expanded (NLM-e). Philos. Trans. Biol. Sci. 363 (1493), 979–1000. https://doi.org/10.1098/rstb.2007.2154.

Kushnerenko, E.V., Van den Bergh, B.R.H., Winkler, I., 2013. Separating acoustic deviance from novelty during the first year of life: a review of event- related potential evidence. Front. Psychol. 4, 595. https://doi.org/10.3389/fpsyg.2013.00595.

Marklund, E., Lacerda, F., Schwarz, I.C., 2018. Using rotated speech to approximate the acoustic mismatch negativity response to speech. Brain Lang. 176, 26–35. https://doi.org/10.1016/j.bandl.2017.10.006.

Maurer, D., Werker, J.F., 2014. Perceptual narrowing during infancy: a comparison of language and faces. Dev. Psychobiol. 56 (2), 154–178. https://doi.org/10.1002/dev.21177.

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.

Molnar, M., Polka, L., Baum, S., Steinhauer, K., 2014. Learning two languages from birth shapes pre-attentive processing of vowel categories: electrophysiological correlates of vowel discrimination in monolinguals and simultaneous bilinguals. Biling. Lang. Cogn. 17 (3), 526–541. https://doi.org/10.1017/S136672891300062X.

Morr, M.L., Shafer, V.L., Kreuzer, J.A., Kurtzberg, D., 2002. Maturation of mismatch negativity in typically developing infants and preschool children. Ear Hear. 23 (2), 118–136. https://doi.org/10.1097/00003446-200204000-00005.

Näätänen, R., Paavilainen, P., Rinne, T., Alho, K., 2007. The mismatch negativity (MMN) in basic research of central auditory processing: a review. Clin. Neurophysiol. 118 (12), 2544–2590. https://doi.org/10.1016/j.clinph.2007.04.026.

Polka, L., Werker, J.F., 1994. Developmental changes in perception of nonnative vowel contrasts. J. Exp. Psychol. Hum. Percept. Perform. 20 (2), 421–435.

Rivera-Gaxiola, M., Silva-Pereyra, J., Kuhl, P.K., 2005. Brain potentials to native and non-native speech contrasts in 7-and 11-month-old American infants. Dev. Sci. 8 (2), 162–172. https://doi.org/10.1111/j.1467-7687.2005.00403.x.

Shafer, V.L., Yu, Y.H., Datta, H., 2011. The development of English vowel perception in monolingual and bilingual infants: neurophysiological correlates. J. Phon. 39 (4), 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? Neurosci. Lett. 526 (1), 10–14. https://doi.org/10.1016/j.neulet.2012.07.064.

Sharma, A., Dorman, M.F., 1999. Cortical auditory evoked potential correlates of categorical perception of voice-onset time. J. Acoust. Soc. Am. 106 (2), 1078–1083. https://doi.org/10.1121/1.428048.

Shibasaki, H., Miyazaki, M., 1992. Event-related potential studies in infants and children. J. Clin. Neurophysiol. 9 (3), 408–418.

Shucard, J.L., Shucard, D.W., 1990. Auditory evoked potentials and hand preference in 6-month-old infants: possible gender-related differences in cerebral organization. Dev. Psychol. 26 (6), 923. https://doi.org/10.1037/0012-1649.26.6.923.

Sjölander, K., Beskow, J., 2000. Wavesurfer - an open source speech tool. Paper Presented at the Sixth International Conference on Spoken Language Processing (ICSLP 2000).

Tiitinen, H., May, P., Reinikainen, K., Näätänen, R., 1994. Attentive novelty detection in humans is governed by pre-attentive sensory memory. Nature 372 (6501), 90–92. https://doi.org/10.1038/372090a0.

Werker, J.F., 1995. Exploring developmental changes in cross-language speech perception. In: In: Gleitman, L.R., Liberman, M. (Eds.), Language: an Invitation to Cognitive Science, vol. 1. pp. 87–106.

Werker, J.F., Curtin, S., 2005. PRIMIR: a developmental framework of infant speech processing. Lang. Dev. 1 (2), 197–234. https://doi.org/10.1207/s15473341lld0102_4.

Werker, J.F., Tees, R.C., 1984. Cross-language speech perception: evidence for perceptual reorganization during the first year of life. Infant Behav. Dev. 7 (1), 49–63. https://doi.org/10.1016/S0163-6383(84)80022-3.

Winkler, I., 2007. Interpreting the mismatch negativity. J. Psychophysiol. 21 (3-4), 147–163. https://doi.org/10.1027/0269-8803.21.34.147.

Winkler, I., Lehtokoski, A., Alku, P., Vainio, M., Czigler, I., Csépe, V., et al., 1999. Pre-attentive detection of vowel contrasts utilizes both phonetic and auditory memory representations. Cogn. Brain Res. 7 (3), 357–369. https://doi.org/10.1016/S0926-6410(98)00039-1.

Xu, D., Yapanel, U., Gray, S., Baer, C.T., 2008. The LENA Language Environment Analysis System: the Interpreted Time Segments (ITS) File. LENA Foundation., Boulder, CO, USA.

Yoshida, K.A., Pons, F., Maye, J., Werker, J.F., 2010. Distributional phonetic learning at 10 months of age. Infancy 15 (4), 420–433. https://doi.org/10.1111/j.1532-7078.2009.00024.x.