As for the chosen time window, the previous literature on infant MMR used various windows for vowels (e.g., 0–500 ms after stimulus onset in Cheour-Luhtanen et al., 1995; 200–500 ms in Cheour et al., 1998a) and various windows for 2- or 3-month olds (e.g., 0–1000 ms in Friederici et al., 2002; 200–600 ms in Friedrich et al., 2004; 100–450 ms and 550–900 ms in He et al., 2009). The only publication on vowels with infants in our age range (3-month olds: Cheour et al., 2002b) used a window from 150 to 400 ms. Regarding the reported variation, and because control of the Type I error rate dictates that analysis windows be chosen before the ERP results are seen, we had to choose in advance a window that includes at least the possible times at which the MMR can occur, namely a window running from 100 to 500 ms. In order to submit this window to an analysis of variance (ANOVA), we divide it into eight consecutive time bins of 50 ms each (Cheour-Luhtanen et al., 1995; Morr et al., 2002; Friedrich et al., 2004; He et al., 2009), and compute the average amplitude of the difference waveform in each bin as our measurement variable. To conclude, each infant's MMR waveform is reduced to only 64 (8 time bins × 8 channels) MMR amplitude values.

STATISTICAL ANALYSIS

To test whether there is a difference between unimodally and bimodally trained infants, while controlling for differences in the presented standard, we subjected the QS and non-QS datasets separately to an ANOVA with a mixed design (between-subject factors and repeated measures). The MMR amplitude was the dependent variable, Time Bin (100–150, 150–200, 200–250, 250–300, 300–350, 350–400, 400–450, and 450–500 ms) and Electrode (Fz, F3, F4, Cz, C3, C4, T7, and T8) were within-subject factors, and Distribution Type (unimodal vs. bimodal) and Standard Vowel ([æ] vs. [ε]) were between-subject factors. The design also included all possible interactions between the factors, up to the fourth order. To compensate for the double chance of finding results (separate QS and non-QS analyses) all tests employ a conservative α level of 0.025.

RESULTS

The grand average waveforms for each Distribution Type (unimodal vs. bimodal) pooled over the two levels of the factor Standard Vowel are presented in **Figure 2**, for 10 electrodes. In line with previous research on 2-to-3-month olds, the standard and deviant ERPs contained prominent slow positive waves (e.g., Friederici et al., 2002; Morr et al., 2002; Carral et al., 2005; Shafer et al., 2011), and the ERPs in the QS data appeared large compared to those in the non-QS data (e.g., for 2-month olds: Friederici et al., 2002; for newborns: Pihko et al., 2004; Sambeth et al., 2009; but see Cheour et al., 2002a, for conflicting results).

For the QS data, the ANOVA on the MMR amplitude yielded significant results neither for the research question (main effect of Distribution Type: $p = 0.88$), nor for any other main effect (Standard Vowel: $p = 0.23$; Electrode: $F < 1$; Time Bin: $F < 1$), nor for any of the 11 interactions (all p-values > 0.07).

For the non-QS data, the ANOVA revealed a positive grand mean (+0.84 μV), with a 97.5% confidence interval (CI) that does not include zero (+0.35 ∼ +1.33 μV), implying that on average Dutch 2-to-3-month old infants can discriminate the test vowels, and that vowel discrimination in these infants is reflected in a *positive* MMR. Regarding our specific research question, the analysis showed a main effect of Distribution Type (mean difference = +1.06 μV, CI = +0.08 ∼ +2.04 μV, $F[1,18] = 7.03$, $p = 0.016$, $\eta_p^2 = 0.28$): across electrodes and time windows the bimodally trained infants had a higher positive MMR (+1.37 μV, CI = +0.68 ∼ +2.06 μV) than the unimodally trained infants (+0.31 μV, CI = –0.38 ∼ +1.00 μV), indicating that Dutch 2-to-3-month olds' neural discrimination of [æ] and [ε] is better after bimodal than after unimodal training.

As for factors not directly pertaining to our research question, there was no effect of Standard Vowel ($p = 0.98$), so that we cannot state with confidence that one of the two combinations of standard and deviant vowel yields a higher MMR amplitude (and thus better neural discrimination) than the other combination. Further, the analysis showed no main effects of Time Bin ($F[7\varepsilon,126\varepsilon$, $\varepsilon = 0.334] = 1.37$, Greenhouse–Geisser corrected $p = 0.27$) or Electrode ($F < 1$). Thus, there was no support for a more positive or more negative MMR in any specific time window as compared to other ones within 100 and 500 ms, and at any specific electrode as compared to other ones among the frontocentral and temporal electrodes. Interestingly, we found a highly significant interaction effect between Distribution Type and Standard Vowel [$F(1,18) = 20.22$, $p = 0.0003$, $\eta_p^2 = 0.53$], which shows that the attested difference between unimodally and bimodally trained Dutch 2-to-3-month olds differs depending on the standard that they hear in the oddball test (see section Exploratory Results for the Four Groups).

EXPLORATORY RESULTS FOR THE FOUR GROUPS

To examine the responses of the four non-QS groups separately, we pooled the MMR amplitudes across electrodes and time bins in view of the lack of significant differences herein (see section Results). **Figure 3** shows the pooled MMR waveforms per group, and **Table 2** lists the corresponding averaged MMR amplitudes. The amplitude differed from zero significantly only for the Bimodal [ε] group ($p = 0.004$, uncorrected for multiple comparisons) implying that bimodally trained Dutch 2-to-3-month olds who are tested with standard [ε] and deviant [æ] can hear the difference between the two vowels.

The individual group's MMR amplitudes presented in **Table 2** are visualized in **Figure 4**. The interaction between Distribution Type and Standard Vowel, which was found in the main ANOVA for the non-QS data (see section Results), is clearly visible. We did the four relevant group comparisons, assuming equal variances for all groups (as in the ANOVA): Bimodal [ε] vs. Unimodal [ε], Bimodal [æ] vs. Unimodal [æ], Bimodal [ε] vs. Bimodal [æ] and Unimodal [æ] vs. Unimodal [ε] (technically, this was done via *post hoc* comparisons using Fisher's Least Significant Difference in SPSS). The Bimodal [ε] group's response was reliably more positive than that of the Unimodal [ε] group (see the arc numbered 1 and the black line in **Figure 4**; uncorrected $p = 0.00008$); this indicates that when the standard in the oddball paradigm is [ε] and the deviant is [æ], bimodally trained Dutch 2-to-3-month olds show better neural discrimination than unimodally trained infants. The difference between Bimodal [æ] and Unimodal [æ] was not significant ($p = 0.21$); thus, when the standard is [æ]