Developmental Science

Developmental Science 15:1 (2012), pp 2–11 DOI: 10.1111/j.1467-7687.2011.01092.x

PAPER

Vowels in early words: an event-related potential study

Nivedita Mani,[1] Debra L. Mills[2] and Kim Plunkett[3]

1. Language Acquisition Junior Research Group, University of Goettingen, Germany
2. School of Psychology, Bangor University, UK
3. Department of Experimental Psychology, University of Oxford, UK

Abstract

Previous behavioural research suggests that infants possess phonologically detailed representations of the vowels and consonants in familiar words. These tasks examine infants' sensitivity to mispronunciations of a target label in the presence of a target and distracter image. Sensitivity to the mispronunciation may, therefore, be contaminated by the degree of mismatch between the distracter label and the heard mispronounced label. Event-related potential (ERP) studies allow investigation of infants' sensitivity to the relationship between a heard label (correct or mispronounced) and the referent alone using single picture trials. ERPs also provide information about the timing of lexico-phonological activation in infant word recognition. The current study examined 14-month-olds' sensitivity to vowel mispronunciations of familiar words using ERP data from single picture trials. Infants were presented with familiar images followed by a correct pronunciation of its label, a vowel mispronunciation or a phonologically unrelated non-word. The results support and extend previous behavioural findings that 14-month-olds are sensitive to mispronunciations of the vowels in familiar words using an ERP task. We suggest that the presence of pictorial context reinforces infants' sensitivity to mispronunciations of words, and that mispronunciation sensitivity may rely on infants accessing the cross-modal associations between word forms and their meanings.

Introduction

The last decade has seen a surge of interest in the accuracy of infants' representations of the sounds of words in accessing word–object associations. Much of this interest originates initially from an habituation-switch study by Stager and Werker (1997) reporting that 14-month-olds cannot simultaneously learn two word–object associations differing by only a single consonant (e.g. *bih*-*dih*). There are two possible conclusions that can be drawn from this finding. The first is that infants do not possess phonologically detailed representations, i.e. their representations of words are not detailed enough to differentiate between *dih* and *bih* (Charles-Luce & Luce, 1990). This conclusion is not the preferred explanation of Stager and Werker (1997) and is challenged by the findings of a number of behavioural studies using the Inter-modal Preferential Looking task (IPL) that find that infants as young as 12 months of age are sensitive to small vowel and consonant mispronunciations of familiar words (12-months: Mani & Plunkett, in press; 14-months: Swingley & Aslin, 2002; Fennell & Werker, 2003; Mani & Plunkett, 2007; 18-months: Bailey & Plunkett, 2002; Mani & Plunkett, 2007; Mani, Coleman & Plunkett, 2008; Mani & Plunkett, 2008a; 18–24 months: Swingley & Aslin, 2000). The alternative conclusion (Stager and Werker, 1997) is that infants may not be able to access this phonological detail in some tasks, perhaps owing to task difficulty, even though their representations of words contain enough phonological detail to discriminate between them.

Stager and Werker (1997) reach their conclusion on the basis of a habituation task. Infants are habituated to a novel word–object association (Object A with the word 'bih') followed by presentation of the same novel object with a switched label, i.e. 'dih'. Fourteen-month-olds failed to notice the switched pairing of Object A with the new label 'dih'. In contrast, IPL studies typically present infants with two images of familiar objects side-by-side on a screen, followed by a correct or mispronunciation of the label for one of the presented images. In this task, infants successfully discriminate correct pronunciations from mispronunciations of the labels for the presented images. Two important differences between the habituation studies and the IPL studies may account for this contrasting pattern of responding. First, Stager and Werker examined infants' sensitivity to changes to newly learnt words, i.e. words to which infants were exposed in a laboratory setting immediately prior to testing. The other studies, however, assess infants' sensitivity to the sounds of familiar words: words infants acquired in a natural setting prior to testing. The contrast between the two studies may, therefore, rest on the assumption that infants pay more attention to the sounds of familiar

Address for correspondence: Nivedita Mani, Language Acquisition Junior Research Group, Gosslerstrasse, 14, 37073 Göttingen, Germany; e-mail: nmani@gwdg.de

words relative to newly learnt words. Evidence against this perspective comes from work suggesting that infants at this age are sensitive to mispronunciations of newly learnt words (Ballem & Plunkett, 2005; Mani & Plunkett, 2008b; Yoshida, Fennell, Swingley & Werker, 2009). A second reason for the contrast may pertain to the cognitive demands of the tasks being presented to infants. For instance, an ERP study shows that, in certain tasks, infants may have difficulty accessing the phonological form of even familiar words (Mills, Prat, Zangl, Stager, Neville & Werker, 2004). Mills *et al.* presented infants with lists of correctly pronounced words, mispronunciations of these words and non-words, and measured the brain potentials associated with the different lists. Mills *et al.* found that, at 14 months of age, there was no difference in the brain potentials to correct pronunciations and mispronunciations although both differed from ERPs to phonetically unrelated nonsense words. Fourteen-month-olds were not sensitive to small changes in the pronunciations of even familiar words. By 20 months of age, infants did discriminate between correct and incorrect pronunciations of the words presented to them in the absence of referential context.[1] Is referential context, however, required for infants to display sensitivity to mispronunciations of familiar words? For instance, in the Mills *et al.* (2004) task, in order for infants to perceive *gare* as a mispronunciation of *bear*, they must first activate the representation of the word *bear* from the mispronunciation. This internal generation of the intended representation of a mispronounced word can be difficult for infants, especially 14-month-olds, who may not have that much familiarity with the word *bear* for it to be activated upon hearing a mispronunciation.

A recent study by Fennell and Waxman (2010) highlights the importance of providing infants with referential context in order to assess their sensitivity to mispronunciations. Fennell and Waxman use an adapted habituation task, where infants were initially exposed to pictures of familiar objects and the labels for these objects (e.g. *shoe*). This was followed by the Stager and Werker switch task (described above). The authors found that 14-month-olds can simultaneously learn two minimally different words, just so long as infants were properly inducted into the referential context of the task, i.e. the task presents infants with objects and the labels for these objects through the prior presentation of familiar object–label pairs. The study highlights the importance of infants' understanding of the referential context of the task, i.e. the association between objects and their labels, in order for them to succeed in paying attention to the phonological detail associated with words. Procedures such as the adapted habituation switch task used by Fennell and Waxman (2010) and the IPL mispronunciation studies may, therefore, provide a cognitively less demanding approach to examining infants' sensitivity to changes to the phonological structure of words than either the Stager and Werker (1997) or the Mills *et al.* (2004) studies.

Nevertheless, there are concerns with the interpretation of the processes underlying infants' sensitivity to mispronunciations in the context of the preferential looking task. The preferential looking task typically presents infants with two images of familiar objects simultaneously followed by either a correct or an incorrect pronunciation of the label for one of the objects. In order to demonstrate robust attention to the phonological detail of the words used in this task, infants need to exhibit diminished preference for the target image and/or reject the distracter image as the intended referent for the mispronunciation. Performance in this task, therefore, depends not just on infants' knowledge of the pronunciation of the label for the target image (i.e. the image whose label has been mispronounced), but also the distracter image. For example, for vowel-medial mispronunciations, an important control is that target-distracter labels begin with the same onset consonant (e.g. *bib*-*bed*: Mani & Plunkett, 2007; Mani *et al.*, 2008). This ensures that infants' looking to target and distracter pictures is influenced by the word-medial vowel (correct or mispronounced) and not by a difference in the onset consonant of the target and distracter labels. This shared onset consonant may also impact the influence of the distracter image (and its label) on infants' responding. For instance, a vowel mispronunciation of *bib* as *beb* may lead to less robust target recognition, when the mispronounced vowel is more similar to the distracter vowel (*bed*) than the target vowel – infants, in this case, would be prompted to look more at the distracter image due to greater overlap between the mispronunciation and the distracter label. Hence, the phonological characteristics of the label associated with the distracter image have the potential to impact infants' sensitivity to mispronunciations of the target label in complex ways. A simpler, less contaminated, estimate of infants' sensitivity to vowel mispronunciations might therefore be more readily achieved in the absence of confounding distracter images, i.e. single picture trials.

In fact, the habituation/switch procedure (Stager & Werker, 1997; Fennell & Waxman, 2010) uses cross-modal single picture presentation. However, this procedure requires pre-habituation to label–object associations, which is time-consuming and potentially compromises the referential character of the task (cf. repetition priming effects in adults). That is, because infants are necessarily habituated to repeated presentations of the same label–object pairs, sensitivity to a change in the acoustic stimulus may rely on more short-term working memory demands in these tasks. Other tasks that rely on infants' sensitivity to mispronunciations of words from single presentations of familiar word–object associations may give us a more

---

[1] Mani and Plunkett (2010) have demonstrated that 18-month-old infants can internally generate the name of an object in a preferential looking task. If 18-month-old infants can also internally generate the mental representation of an object when they hear its name, then they may also be able to provide their own referential context.

accurate index of the phonological detail associated with infants' representations of words. Furthermore, the habituation/switch procedure has only been used successfully with infants under 20 months of age and, therefore, undermines the potential for direct comparisons with older infants.

Event-related potentials (ERPs), on the other hand, may provide an index of the suitability of a heard word as the label for a presented image using single picture trials. ERPs are averaged epochs of electrical activity time-locked to a particular stimulus event, such as the presentation of a label for a single displayed image. Previous ERP studies using a cross-modal paradigm to investigate brain activity to mispronounced words have identified two electrophysiological indices of phonological and semantic processing in adults. The first is an early negativity between 250 to 350 ms after stimulus onset, sensitive to changes to the expected phonological form of a word, sometimes referred to as a Phonological Mismatch Negativity (PMN) (Connolly & Phillips, 1994; Newman, Connolly, Service & McIvor, 2003; D'Arcy, Connolly & Crocker, 2000). Using a picture–word matching paradigm with adults similar to the current study, Desroches, Newman and Joanisse (2008) show an early negative component sensitive to consonant changes (e.g. *bone*, *comb*) to the labels of visually presented images (*cone*). In addition, the consonant changes also influenced a later negative component (the N400), with a more negative going wave for single consonant changes (e.g. *bone*, *comb*) and completely unrelated words (e.g. *fox*) than for matching labels (*cone*). The N400 is an index of semantic processing, indicating the integration of a stimulus into prior semantic context (Kutas & Hillyard, 1983). Changing the way a word is pronounced influences the meaning of the word, resulting in disruption of the ease of integrating the image (*cone*) with the auditory stimulus (*comb*), as reflected by the N400.

Previous research shows a similar modulation of infants' brain potentials in these time windows using picture–word matching tasks (Sheehan, Namy & Mills, 2007; Friedrich & Friederici, 2004). Sheehan *et al.* report finding an early (200–400 ms) and later negative component (400–600 ms) influenced by the congruence of an auditorally presented word as a label for a visually presented image, with a more negative going wave for incongruous word–image pairings (e.g. *cup-book*) compared to congruous pairings (e.g. *cup-cup*) at 18 and 26 months of age. The 200–400 ms time window also showed significant differences between mispronunciations and correct pronunciations in 20-month-olds in the Mills *et al.* study (2004). Similarly, Friederich and Friederici (2004) find early effects (between 150 to 400 ms) of semantic congruence of picture–word pairings (e.g. apple-apple vs. apple-book) in 19-month-olds, with more negative responses to incongruous words than congruous (see also Torkildsen, Sannerud, Syversen, Thormodsen, Simonsen, Moen, Smith & Lindgren, 2006; Mills, Conboy & Paton, 2005, for other studies reporting modulation of the N400 component by semantic congruency).

The current study, therefore, examines infants' sensitivity to vowel mispronunciations of familiar words by analysing the ERPs to single picture trials, with the aim of obtaining an estimate of infants' attention to vocalic detail in the absence of confounding distracter images. The age tested in the current study, i.e. 14 months, provides a useful comparison with previous work on the specificity of infants' phonological representations, given that both the Mills *et al.* (2004) study and Stager and Werker (1997) test infants at this age. However, we use a contrasting method of stimulus presentation and focus on vowel mispronunciations as opposed to consonant mispronunciations. Based on Fennell and Waxman's (2010) demonstration of the importance of infants' understanding of the referential context of the task for attending to phonological detail, we present infants with cross-modal stimuli, i.e. a visual image and an auditory label for this image. We believe that the concurrent presentation of image and label will help infants better assess the quality of this auditory label relative to their stored representation of the label for this image, and allow us to estimate infants' attention to the phonological make-up of the auditory label in the presence of pictorial context. The ERP approach also has the advantage of offering fast brain responses and high levels of temporal resolution that are advantageous for investigation of speech processing. Using a cross-modal design, we examine whether the ERP components to correct and incorrect pronunciations vary in latency or amplitude, such that ERPs to mispronunciations are larger in amplitude than to correct pronunciations.

Based on the results of previous ERP studies, we focus our analysis on the time windows 200 to 600 ms after onset of the label (correct or incorrect). If 14-month-olds display sensitivity to mispronunciations of the labels for the images presented to them, we would expect to find a significant N400 effect, and possibly an earlier effect around 200 ms (N200–300). We also expect a significant N400, and possibly an early phonological negativity to a control condition, where the auditory stimulus is a non-word phonetically dissimilar in its entirety to the label for the presented image. Note, however, that it is also possible that if the PMN and the N400 tap separably into pre-lexical and lexical stages of processing, we may find *only* a PMN; suggesting that infants are sensitive to mispronunciations of words for which they do not as yet have fully formed semantic representations (as has been shown in behavioural studies; Swingley, 2003). Comparing the pattern of results across the PMN and N400 may, therefore, provide us with an assessment of the level at which the mispronunciations are being detected, i.e. pre-lexical or lexical. The control condition also allows examination of whether the ERPs to mispronunciations pattern with these phonetically dissimilar non-words or with correct pronunciations, providing an indication of the lexical status of mispronunciations. The absence of a significant difference between non-words and mispronunciations would indicate that the non-words and mispronunciations are both incongruous to the displayed image, while the correct pronunciations are more easily integrated with the displayed image. Furthermore, given the evidence in favour of left-hemisphere specialization of the response to mispronunciations reported by Mills *et al*. (2004) and the response to familiar versus unknown words in Mills, Coffey-Corina and Neville (1993, 1997), we also examine the pattern of sensitivity to mispronunciations and non-words separately in the left and right hemispheres. In keeping with these results, we may find that at 14 months, any significant differences across conditions are broadly distributed across the scalp, since Mills and colleagues find left-hemisphere specialization of the response to mispronunciations only later, at 20 months of age. However, the presentation of an image may help to create a referential context for the label, leading to specialization of the response to a mispronunciation even in young infants.

Method

Participants

The participants in this experiment were 16 infants at 14 months of age (M = 14.4, range = 13.9 to 14.8). Five additional infants were tested but were excluded due to their failure to provide at least 10 artifact-free trials per condition. All infants had no known hearing or visual problems, were born full-term and were recruited via the local maternity ward. Infants came from homes where British English was the only language in use. Parents gave their informed consent and were given a free T-shirt and/or travel expenses as compensation for participation. Prior to beginning the study, parents were asked to indicate whether their children understood the meanings of the words presented to them in the study by completing the British Communicative Development Inventory (Hamilton, Plunkett & Schafer, 2000, an adaptation of the MacArthur-Bates CDI; Fenson, Dale, Reznick, Thal, Bates, Hartung, Pethic & Reilly, 1993).

Stimuli

The speech stimuli were produced by a female speaker of British English in an enthusiastic, child-directed manner. The audio recordings were made with a solid state compact flash card recorder in a sound-treated recording booth. The audio stimuli were digitized at a sampling rate of 44.1 kHz and a resolution of 16 bits and spliced using Goldwave v. 5.10. The stimuli presented to children were 30 monosyllabic nouns taken from the British Communicative Developmental Inventory (Hamilton *et al*., 2000). Mispronunciations changed the word-medial vowel of these words by vowel height (e.g. image of a bed along with the label *bid*) or vowel backness (e.g. image of a brush along with the label *brash*) to another standard English vowel. In addition, we also presented infants with 30 phonotactically legal non-words. Non-words were created by combining the onset consonant of one of the familiar words with the medial vowel and final consonant of another word, e.g. such a combination of *milk* and *dog* yields the non-word *mog*. Non-words were paired with images whose labels were phonetically completely dissimilar to the non-word, e.g. *mog* presented with an image of a *cup*. We ensured that there were no systematic differences in the duration of the correct pronunciations (M = 468 ms, SE = 15) and non-words (M = 480 ms, SE = 15; p = .3). We also attempted to equate vowel durations for correct pronunciations (M = 198 ms; SE = 12) and mispronunciations (M = 212 ms; SE = 12) by asking the speaker to pronounce the word in a particular way ($p > .4$). A complete list of the auditory stimuli used in the experiment is provided in the Appendix.

Visual stimuli were computer images created from photographs, with three different images for each familiar word. All subjects saw all three images for the familiar word paired with a different condition, i.e. correct pronunciations of the label for the image, mispronunciations, or non-words. Familiar images were judged by three adults (two of the authors and an independent observer) as typical exemplars of the labelled category.

Procedure

Stimulus presentation

After the electrode cap placement, children sat on their caregiver's lap during the experiment 80 cm away from a projection screen. Auditory stimuli were presented through a centrally located loudspeaker located immediately above the screen at an average of 65 dB. Visual stimuli (measuring 30 cm × 24 cm) were centrally located on the presentation screen (55 cm in diameter).

Each infant was presented with 90 trials; 30 correct pronunciation trials, 30 vowel mispronunciation trials and 30 non-word trials. Each trial began with the presentation of an animated cartoon dancing to music on the middle of the screen until infants fixated the centre of the screen. To ensure that infants attended to the screen, a researcher sat in the testing booth and pressed a button to start each trial when the infant oriented to the screen. Once the researcher determined that infants were paying attention, they were presented with a centrally located image of a familiar object. This image remained on screen in silence for a predetermined period, at the end of which infants heard the label for the displayed image. The heard label would be either a correct pronunciation, a vowel mispronunciation of the label for the image, or a phonetically dissimilar non-word. The onset of the label varied from 750 to 1250 ms, i.e. at 750 ms, 850 ms, 950 ms, 1050 ms, 1150 ms, 1250 ms. The timing of presentation of the label for the image varied across trials in order to ensure that infants did not begin to expect the label to be presented at a certain time in each trial. Varying the SOA of the visual and auditory stimuli ensures that subjects do not consciously or unconsciously predict the onset of the auditory stimuli leading to brain potentials yoked to the expected timing of the stimulus. The timing of the onset of the label was counterbalanced across conditions across infants. The image remained on-screen for 1500 ms after the onset of the label.

Infants saw three different images for each familiar object during the course of the experiment. Each *image* was, therefore, presented only once during the experiment paired with a correct pronunciation, a vowel mispronunciation or a non-word. The pairing of pronunciation condition and each of the three images for the familiar object was counterbalanced across infants, ensuring that across infants we compare the same visual stimuli across conditions. Infants, therefore, never heard the same word repeated during the experiment. The order of presentation of the three pronunciation conditions during the experiment was interleaved but counterbalanced across words and across infants, ensuring a gap of at least 15 trials between different pronunciations of the same word. Therefore, a third of the infants received the correct pronunciation of a word first, a third received the mispronunciation first, and a third received the non-word first.

Event-related potential recording

EEGs were recorded continuously from tin electrodes at 19 channels attached to a stretchable electrode cap, with two additional electrodes placed on the mastoids and one electrode located above the eye. Electrodes were placed in the standard 10/20 locations, i.e. FP1/FP2, F7/F8, F3/F4, T3/T4, C3/C4, T5/T6, P3/P4, O1/O2, FZ, CZ, PZ, and A1/A2. See Figure 1 for a visual display of electrode site locations on the electrode cap. EEG was referenced online to A1 and re-referenced offline to averaged mastoids (A1/A2). EEG was digitized at 1000 Hz, with a band-pass filter of 0.1 Hz to 70 Hz. All impedances were maintained below 10 kΩ.

Event-related potential analysis

Averaging and artifact rejection was carried out offline using Neuroscan analysis software (Scan 4.3). Artifact rejection thresholds were calculated individually for each infant after inspection of the infant's blinking and eye-movement data. Infants were excluded from the analysis if they provided fewer than 10 trials per condition. In addition, those trials where parents indicated that their infants did not know the words presented to them were also excluded from analysis. This resulted in the exclusion of 33.6% of trials, with a mean of 19.58 trials included per condition (range = 12.0 to 24.6). The EEG data were time-locked to the onset of the auditory stimulus using epochs from 100 ms prior to word onset until 1000 ms after word onset. The data from individual electrodes were then corrected to the 100 ms pre-stimulus baseline and averaged according to condition, i.e. correct pronunciations, mispronunciations and non-words. A 30 Hz low-pass filter was applied to the data post-averaging. We then analysed the data in 50 ms time windows (from 0 ms to 1000 ms) to determine the onset and offset of significant differences between conditions. Based on this analysis, and adult and infant studies showing functionally distinct components for phonological and semantic processing (e.g. Connolly & Phillips, 1994; Sheehan *et al.*, 2007), we focused our analyses on two separate time windows; between 200 ms to 300 ms and between 400 ms to 600 ms (the N400 window). For purposes of data reduction, a selection of electrode locations was entered into data analysis, divided into four regions from front to back of the head: fronto-central (F3/F4), fronto-lateral (F7/F8), anterior-temporal (T3/T4) and central (C3/C4). The central

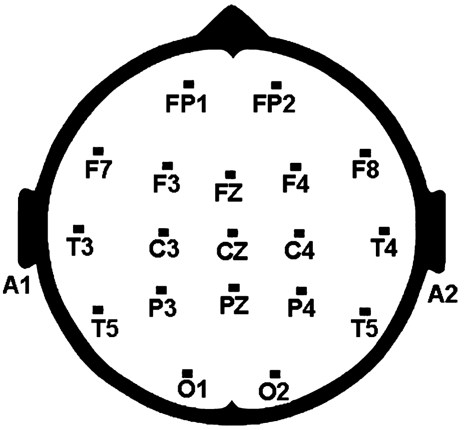

Figure 1 *Placement of electrodes in International 10–20 array. Results are reported for F7, F8, F3, F4, T3, T4, C3 and C4.*

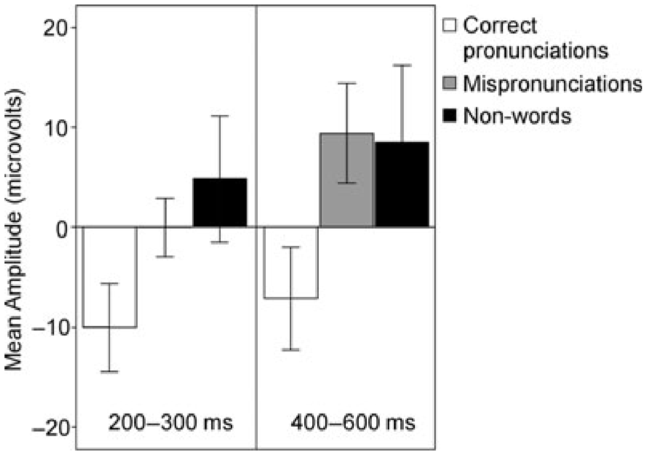

Figure 2 *ERPs to mispronunciations and non-words were larger (more negative going) than to words in both the 200–300 and 400–600 time windows over the left-hemisphere sites F7, F3, T3, and C3.*

electrodes were also entered into a separate analysis to check for consistency with the lateral sites. This analysis examined whether there was a difference in the amplitude and latency of the brain potentials following correct pronunciations, mispronunciations and non-words. Effect sizes are reported using partial eta squared. For purposes of clarity, only main effects of condition and interactions with condition will be reported. Because our hypotheses pertain to planned comparisons between specific conditions rather than a condition effect in general, we do not report the omnibus ANOVA (see Ableson & Prentice, 1997).

Results

200–300 ms

We examined the difference in the mean amplitude of the brain potentials following correct pronunciations, mispronunciations and non-words in this 200 ms to 300 ms window following auditory stimulus onset (see Figures 2 and 3). A repeated measures ANOVA comparing correct pronunciations and non-words found a near-significant interaction between condition, hemisphere and electrode site ($F(3, 13) = 3.20$, $p = .059$, $\eta_p^2 = .42$). Although we didn't find a similar interaction comparing correct pronunciations and mispronunciations (ps > .1), based on visual inspection of the individual data and previous work showing hemisphere effects in similar paradigms (Mills *et al.*, 2004), planned comparisons examined the effects across the three conditions separately for the left and right hemispheres.

Left hemisphere

There was a significant difference between correct pronunciations and non-words ($F(1, 15) = 4.65$; $p = .04$, $\eta_p^2 = .23$), with the N200–300 to non-words being more negative than to correct pronunciations. Similarly, there was a significant difference between correct and mispronunciations, with the N200–300 to mispronunciations being more negative than to correct pronunciations ($F(1, 15) = 7.53$; $p = .015$; $\eta_p^2 = .33$). This effect was found in left fronto-central and temporal regions. Thirteen out of 16 participants showed this N200–300 effect for both mispronunciations and non-words. There was no difference in the N200–300 to mispronunciations and non-words ($F(1, 15) = .35$; $p = .5$).

Right hemisphere

There was no significant difference between correct pronunciations and non-words ($F(1, 15) = .6$; $p = .4$), between correct and mispronunciations ($F(1, 15) = .7$; $p = .4$) or between mispronunciations and non-words ($F(1, 15) = .4$; $p = .5$).

Overall, these results suggest that while infants were sensitive to the pairing of a mispronunciation or a phonetically dissimilar word with the familiar image, infant sensitivity to both mispronunciations and phonetically dissimilar non-words was more prominent in the left than in the right hemisphere.

400–600 ms

We examined the difference in mean amplitude of the brain potentials following correct pronunciations, mispronunciations and non-words in the 400 ms to 600 ms window (see Figures 2 and 3). A repeated measures ANOVA comparing correct pronunciations and non-words found a significant interaction between condition, hemisphere and electrode site ($F(3, 13) = 5.66$, $p = .011$, $\eta_p^2 = .56$). A repeated measures ANOVA comparing correct pronunciations and mispronunciations found a significant main effect of condition ($F(1, 15) = 8.15$, $p = .012$; $\eta_p^2 = .35$), but no interactions with condition. As with the earlier time window, based on these analyses and visual inspection of the individual data, planned comparisons examined the effects across the three conditions separately for the left and right hemispheres.

Left hemisphere

There was a significant difference in the N400 to correct pronunciations and non-words, with the N400 to non-words being larger than to correct pronunciations ($F(1, 15) = 4.95$; $p = .04$, $\eta_p^2 = .24$). Similarly, the N400 to correct pronunciations was smaller than to

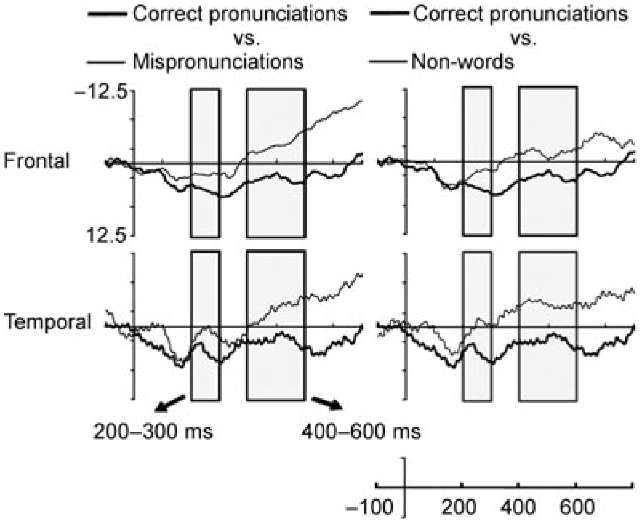

Figure 3 *ERP differences are directly compared to correct pronunciations and mispronunciations (left side) and correct pronunciations and non-words (right side). Significant differences in the N200–300 and N400–600 mean amplitudes are shaded and enclosed in the rectangle. ERPs plotted for F7 and T3. ERPs plotted from −100 ms to 800 ms from word onset.*

mispronunciations (*F*(1, 15) = 11.49; $p = .004$, $\eta_p^2 = .43$). This effect was found in left fronto-central and temporal regions. Twelve out of 16 participants showed this N400 effect for both mispronunciations and non-words compared to correct pronunciations. There was no difference in the N400 to mispronunciations and non-words (*F*(1, 15) = .07; $p = .78$).

Right hemisphere

There was no significant difference in the N400 between correct pronunciations and non-words (*F*(1, 15) = .4; $p = .5$), between correct pronunciations and mispronunciations (*F*(1, 15) = 1.29; $p = .2$) or between mispronunciations and non-words (*F*(1, 15) = .04; $p = .8$).

Discussion

The current experiment aimed to provide an electrophysiological index of infants' sensitivity to mispronunciations of the vowels in familiar words at 14 months of age. Infants were presented with a single image of a familiar object followed by a label for this object. The label was either correctly pronounced, mispronounced by a single vocalic feature, or a phonotactically legal non-word, which was phonologically unrelated to the label for the image. There were significant differences in the brain potentials to the auditory stimuli across two time windows, i.e. 200–300 ms and 400–600 ms after the onset of the auditory stimuli.

During the N200–300 time window, we found a significant difference between correct pronunciation and non-word trials, correct pronunciation and vowel mispronunciation trials, but no significant difference between vowel mispronunciation and non-word trials. Similarly, there was a significant difference between correct pronunciation and non-word trials, and between correct pronunciation and vowel mispronunciation trials across the N400–600 time window. Again, there was no difference between non-word and vowel mispronunciation trials in this time window. These effects were more prominent in the left fronto-central and left temporal regions potentially indexing the violation of the top-down phonological expectations raised by the visual stimulus.

The differences in ERPs to correct pronunciation, mispronunciation and non-word trials can be taken to indicate that, in the context of a familiar image, infants are able to detect mispronunciations of the label for this image by 14 months of age. We interpret this difference as indexing the difficulty of integrating a mispronunciation or a non-word with the phonological expectation raised by the visually presented object. In keeping with previous behavioural results (Mani & Plunkett, 2007, 2008b; Curtin, Fennell & Escudero, 2009), the current study provides the first ERP evidence that infants possess phonetically detailed representations of the vowels in familiar words, and that infants are able to access these phonetically detailed representations in cross-modal, word recognition tasks by 14 months of age.

Experiments to date using eye-fixation methods report that infants look less at an image of a familiar object when the label for this object is mispronounced by either a vocalic or a consonantal feature change. However, most of this work examines infants' sensitivity to the mispronunciation of a label of one of two simultaneously presented images, e.g. comparing infants' looking time at a picture of a bib, relative to a bed, when presented with either a correct or incorrect pronunciation of the word *bib*. As argued previously, in examining infants' sensitivity to vowel mispronunciations, especially, such an approach may be confounded by the similarity of the mispronunciation (e.g. *bib* mispronounced as *beb*) to the label for the distracter image.[2] ERPs, on the other hand, allow evaluation of the congruence of a label (correctly or incorrectly pronounced) to a single image, thereby examining infants' sensitivity to vowel mispronunciations in the absence of confounding distracter images. The current experiment demonstrates infant sensitivity to single feature mispronunciations of the medial vowel of familiar, monosyllabic words, even in the absence of a distracter image.

It is also important to highlight the parallels and contrasts between the results of the current study and Mills *et al.* (2004). Mills *et al.* report not finding a significant difference in the N200–400 effect to consonant mispronunciations and correct pronunciations at 14 months of age, the age group tested in the current study. In contrast, the current study finds robust sensitivity to vowel mispronunciations at the same age. Similar to Mills *et al.* (2004) and previous work by Mills and colleagues, however, is the finding that the effects were predominantly limited to the left hemisphere. Nevertheless, Mills *et al.* (2004) report finding the effects broadly distributed over the scalp at 14 months of age, and limited primarily to left temporal and parietal sites at 20 months of age.

One reason for these differences may be that 14-month-olds are not sensitive to consonant mispronunciations, as tested in Mills *et al.* (2004), but are sensitive to vowel mispronunciations, as tested in the current study. The acoustic cues that differentiate vowels versus consonants may affect the outcomes of studies, particularly in young infants whose lexical representations are still highly flexible. Therefore, future studies should investigate whether the addition of referential context (by presenting infants with an image whose label is mispronounced) influences infants' sensitivity to consonant mispronunciations, in comparison to Mills *et al.* (2004).

---

[2] Evaluation of infants' sensitivity to consonant mispronunciations may not be equally susceptible to a confounding influence from the distracter image, because the mispronounced consonant is always as different from the onset consonant of the distracter label as it is from the target label in the Mani and Plunkett task (where target-distracter labels begin with the same onset consonant).

However, it is unlikely that the difference between the results reported in the current study and that reported by Mills *et al.* is due to any differences in infants' sensitivity to vowel versus consonant mispronunciations. This conjecture is motivated by previous studies (apart from Mills *et al.*, 2004) which find, without exception, that infants at 11 months (Swingley, 2005), 12 months (Mani & Plunkett, in press), 14 months (Swingley & Aslin, 2002; Ballem & Plunkett, 2005; Mani & Plunkett, 2007), 15 months (Mani & Plunkett, 2008b) and 18 months of age (Bailey & Plunkett, 2002; Swingley & Aslin, 2000; Mani & Plunkett, 2007, 2008b) show a robust sensitivity to consonant mispronunciations of familiar words.

A second possibility relates to the finding of individual differences in infants' sensitivity to mispronunciations. For instance, Friederich and Friederici (2006) find differences in the size of the N400 effect displayed by infants with varying degrees of expressive language skills. Since ERPs appear sensitive to such individual differences, it is possible that the contrast between Mills *et al.* (2004) and the current study is related to differences in the expressive language skills of the infants tested. Whilst it is possible that individual differences in infants' ability to detect mispronunciations underlie the contrasting results of the two studies, we focus here on a potentially more revealing difference; i.e. the task used to examine infants' sensitivity.

Indeed, the most plausible alternative explanation for the difference in results is the difference in design between Mills *et al.* (2004) and the current study. The current study presents infants with pictorial context in the form of an image of the familiar object, prior to testing infants' sensitivity to mispronunciations of the label for this image, while Mills *et al.* tested infants' sensitivity to mispronunciations of words in the absence of such pictorial support. Presentation of the image of the familiar object may help infants more readily access the phonological representation of the label for this image and better detect small variations from this stored representation. Pictorial context might, therefore, be crucial to examining infant sensitivity to small variations from the stored phonological representations of familiar words.[3]

It is also noteworthy that, similar to other studies, using prior pictorial (Sheehan *et al.*, 2007; Desroches *et al.*, 2008) or semantic context (van den Brink, Brown & Hagoort, 2001), we find an effect of congruency in an early (200–300 ms) and a later time window (400–600 ms). One interpretation of this finding is that these differences index two potentially separable processes (Connolly & Phillips, 1994; Desroches *et al.*, 2008; van den Brink *et al.*, 2001). The early effects may reflect 'the interface between lexical form and contextual meaning', where word candidates with shared phonological or lexical form are all assessed for their goodness-of-fit to the semantic context (provided by the image in the current study), similar to the PMN (Phonological Mismatch Negativity) components reported by Connolly and Phillips (1994). Indeed, the fronto-central distribution of the N200–300 found in the current study matches the effect reported in van den Brink *et al.* (2001).[4] The later effect, the N400, may reflect higher-order semantic integration of the selected candidate into the context provided.

While these seem potentially viable explanations for the processes underlying the reported effects, it is worth noting the similarity of the distribution and pattern of differences found between 200–300 and 400–600 ms in the current study. Unlike van den Brink *et al.* (2001), both the early (N200–300) and later (N400) congruency effects found in the current study were located at fronto-central and left temporal regions. Furthermore, the differences in the pattern of wave-forms for correct and incongruent labels were similar across the two time windows, suggesting that these two windows may index similar cognitive processes, i.e. infants' assessment of the goodness-of-fit of the mispronunciation and non-word relative to the correct pronunciation as a potential label for the visually presented image.

A similar result is reported by Sheehan *et al.* (2007), with an N400-like congruency effect being found at both 200–400 and 400–600 ms time windows. Also, Friedrich and Friederici (2004) find an early effect of image–label congruence in 19-month-olds between 150 and 400 ms. Friederich and Friederici collapse this time-range as one N400, rather than separate it out into two different effects (as has been suggested in the adult literature reviewed above). Therefore, at least in the context of infant studies (see also Friederich & Friederici, 2005), these early effects (i.e. between 150 and 400 ms) are generally taken to reflect an early onset of the N400. However, we acknowledge that a proper interpretation of the difference between the two time windows is controversial and may require more research into the development of such effects in infancy. For example, in keeping with van den Brink *et al.* (2001), this early time window may reflect infants' processing of phonological form correspondences between the different pronunciation

---

[3] Note that Swingley (2005) finds that 11-month-olds are sensitive to consonant mispronunciations of familiar words even in the absence of pictorial context (in a preferential listening task). However, in keeping with the Stager and Werker (1997) hypothesis, it is possible that 11-month-olds may be displaying sensitivity to changes in phonological form alone, while the 14-month-olds in Mills *et al.* (2004) may have attempted to access more than just phonological form (i.e. correspondences between sequences of phonemes), but also relate this form to word-meaning. This additional load on the 14-month-olds might render them insensitive to the changes they were sensitive to at 11 months.

[4] One might wonder why this component does not pick up on the form-similarity between the mispronunciation and correct pronunciation. The reason for this might lie in the fact that the mispronunciation differs from the correct pronunciation in the vowel, while in the comparable condition in van den Brink *et al.* (2001) the onset matching condition shared the entire first syllable with the target word (*penseel-pension*), not just the onset consonant as in the current study. A similar early incongruency effect has also been reported recently by Kovíc, Plunkett and Westermann (2010) in an ERP study of vowel-driven sound symbolism effects in adulthood.

conditions presented to them, and the later time window indexes the integration of the auditory stimulus into the higher-order semantic context provided by the image.

Irrespective of the correct interpretation, i.e. arguing for either the separability or similarity of the two components, importantly, they both lead to the same conclusion regarding the current study – by 14 months of age, infants display a robust sensitivity to vowel mispronunciations of words. The current study provides the first ERP study of infant sensitivity to vowel mispronunciations of words. Not only does this extend previous behavioural assessments with an electrophysiological correlate of the phonological specificity of infant lexical representations, the current study also establishes the importance of providing infants with a referential context in testing the degree of specificity associated with infants' representations of words.

Acknowledgements

This research was funded by a Wellcome Trust grant awarded to Kim Plunkett. Further analysis and write-up of this research was funded by the German Initiative of Excellence (Nivedita Mani).

References

Abelson, R.P., & Prentice, D.A. (1997). Contrast tests of interaction hypotheses. *Psychological Methods*, **2**, 315–328.

Bailey, T., & Plunkett, K. (2002). Phonological specificity in early words. *Cognitive Development*, **17**, 1265–1282.

Ballem, K., & Plunkett, K. (2005). Phonological specificity in 14-month-olds. *Journal of Child Language*, **32**, 159–173.

Charles-Luce, J., & Luce, P.A. (1990). Similarity neighborhoods of words in young children's lexicons. *Journal of Child Language*, **17**, 205–215.

Connolly, J.F., & Phillips, N.A. (1994). Event-related potential components reflect phonological and semantic processing of the terminal word of spoken sentences. *Journal of Cognitive Neuroscience*, **6**, 256–266.

Curtin, S., Fennell, C., & Escudero, P. (2009). Weighting of acoustic cues explains patterns of word–object associative learning. *Developmental Science*, **12**, 725–731.

D'Arcy, R.C.N., Connolly, J.F., & Crocker, S.F. (2000). Latency shifts in the N2b component track phonological deviations in spoken words. *Clinical Neurophysiology*, **111** (1), 40–44.

Desroches, A.S., Newman, R.L., & Joanisse, M.R. (2008). Investigating the time-course of spoken word recognition: electrophysiological evidence for the influences of phonological similarity. *Journal of Cognitive Neuroscience*, **21**, 1893–1906.

Fennell, C., & Waxman, S.R. (2010). What paradox? Referential cues allow for infant use of phonetic detail in word learning. *Child Development*, **81**, 1376–1383.

Fennell, C.T., & Werker, J.F. (2003). Early word learners' ability to access phonetic detail in well-known words. *Language and Speech*, **46**, 245–264.

Fenson, L., Dale, P.S., Reznick, J.S., Thal, D., Bates, E., Hartung, J.P., Pethic, S., & Reilly, J. (1993). *MacArthur-Bates Communicative Development Inventory: Users' guide and technical manual*. San Diego, CA: Singular.

Friedrich, M., & Friederici, A.D. (2004). N400-like semantic incongruity effect in 19-month-olds: processing known words in picture contexts. *Journal of Cognitive Neuroscience*, **16**, 1465–1477.

Friedrich, M., & Friederici, A.D. (2005). Lexical priming and semantic integration reflected in the event-related potential of 14-month-olds. *NeuroReport, For Rapid Communication of Neuroscience Research*, **16**, 653–656.

Friedrich, M., & Friederici, A.D. (2006). Early N400 development and later language acquisition. *Psychophysiology*, **43** (1), 1–12.

Hamilton, A., Plunkett, K., & Schafer, G. (2000). Infant vocabulary development assessed with a British communicative development inventory. *Journal of Child Language*, **27**, 689–705.

Ković, V., Plunkett, K., & Westermann, G. (2010). The shape of words in the brain. *Cognition*, **114**, 19–28.

Kutas, M., & Hillyard, S.A. (2003). Event-related brain potentials to grammatical errors and semantic anomalies. *Memory and Cognition*, **11** (5), 539–550.

Mani, N., Coleman, J., & Plunkett, K. (2008). Phonological specificity of vocalic features at 18-months. *Language and Speech*, **51**, 3–21.

Mani, N., & Plunkett, K. (2007). Phonological specificity of vowels and consonants in early lexical representations. *Journal of Memory and Language*, **57**, 252–272.

Mani, N., & Plunkett, K. (2008a). Vowels and consonants, difference or no difference. Paper presented to the International Conference on Infant Studies, Vancouver, Canada.

Mani, N., & Plunkett, K. (2008b). Fourteen-month-olds pay attention to vowels in novel words. *Developmental Science*, **11**, 53–59.

Mani, N., & Plunkett, K. (2010). In the infant's mind's ear: evidence for implicit naming in 18-month-olds. *Psychological Science*, **21**, 908–913.

Mani, N., & Plunkett, K. (in press). Twelve-month-olds know their cups from their keps and tups. *Infancy.*

Mills, D., Coffey-Corina, S., & Neville, H. (1993). Language acquisition and cerebral specialization in 20-month-old infants. *Journal of Cognitive Neuroscience*, **5**, 317–334.

Mills, D.L., Coffey-Corina, S., & Neville, H. (1997). Language comprehension and cerebral specialization from 13 to 20 months. *Developmental Neuropsychology*, **13**, 397–445.

Mills, D.L., Conboy, B.T., & Paton, C. (2005). How learning new words shapes the organization of the infant brain. In L.L. Namy (Ed.), *Symbol use and symbolic representation* (pp. 123–153). Mahwah, NJ: Lawrence Erlbaum Associates.

Mills, D.L., Prat, C., Zangl, R., Stager, C.L., Neville, H.J., & Werker, J.F. (2004). Language experience and the organization of brain activity to phonetically similar words: ERP evidence from 14- and 20-month-olds. *Journal of Cognitive Neuroscience*, **16**, 1–13.

Newman, R.L., Connolly, J.F., Service, E., & McIvor, K. (2003). Influence of phonological expectations during a phoneme deletion task: evidence from event-related brain potentials. *Psychophysiology*, **40**, 640–647.

Sheehan, E.A., Namy, L.L., & Mills, D.L. (2007). Developmental changes in neural activity to familiar words and gestures. *Brain and Language*, **101**, 246–259.

Stager, C.L., & Werker, J.F. (1997). Children listen for more phonetic detail in speech perception than word learning tasks. *Nature*, **388**, 381–382.

Swingley, D. (2003). Phonetic detail in the developing lexicon. *Language and Speech*, **46**, 265–294.

Swingley, D. (2005). Eleven-month-olds' knowledge of how familiar words sound. *Developmental Science*, **8**, 432–443.

Swingley, D., & Aslin, R.N. (2000). Spoken word recognition and lexical representation in very young children. *Cognition*, **76**, 147–166.

Swingley, D., & Aslin, R.N. (2002). Lexical neighbourhoods and the word form representations of 14-month-olds. *Psychological Science*, **13**, 480–484.

Torkildsen, J.V.K., Sannerud, T., Syversen, G., Thormodsen, R., Simonsen, H.G., Moen, I., Smith, L., & Lindgren, M. (2006). Semantic organization of basic level words in 20-month-olds: an ERP study. *Journal of Neurolinguistics*, **19**, 431–454.

van den Brink, D., Brown, C.M., & Hagoort, P. (2001). Electrophysiological evidence for early contextual influences during spoken-word recognition: N200 versus N400 effects. *Journal of Cognitive Neuroscience*, **13**, 967–985.

Yoshida, K., Fennell, C.T., Swingley, D., & Werker, J.F. (2009). Fourteen-month-old infants learn similar sounding words. *Developmental Science*, **12**, 412–418.

Received: 10 May 2010
Accepted: 12 June 2011

Appendix

	Correct Pronunciation	Mispronunciation	Non-word
1.	BALL /Bɔ:L/	BU:L	DʊK
2.	BATH /Bɑ:θ/	BÆθ	KU:S
3.	BED /BεD/	BɪD	Kʌʃ
4.	BIB /BɪB/	BεB	DÆT
5.	BIN /BɪN/	BεN	DɪB
6.	BOOK /BʊK/	BɒK	Gɑ:θ
7.	BOOT /BU:T/	Bɔ:T	ʃɪLK
8.	BREAD /BRεD/	BRÆD	SPɒL
9.	BRUSH /BRʌʃ/	BRÆʃ	SPU:T
10.	BUS /BʌS/	BÆS	MɪG
11.	CAT /KÆT/	KεT	Sɪʃ
12.	CHEESE /TʃI:Z/	TʃɪZ	TʌP
13.	CUP /KʌP/	KεP	PɒK
14.	DOG /DɒG/	DʊG	BI:θ
15.	DOLL /DɒL/	DʊL	KÆND
16.	DUCK /DʌK/	DÆK	GɪN
17.	FISH /Fɪʃ/	Fεʃ	TÆT
18.	FOOT /FʊT/	FɒT	DʌS
19.	HAND /HÆND/	HεND	MI:P
20.	HAT /HÆT/	HεT	Gɑ:θ
21.	HEAD /HεD/	HʌD	ʃεN
22.	JUICE /DʒU:S/	Jɔ:S	GU:N
23.	KEYS /KI:Z/	KɪZ	NεD
24.	MILK /MɪLK/	MεLK	SʊT
25.	PEN /PεN/	PʌN	HɒG
26.	PIG /PɪG/	PεG	MI:Z
27.	SHEEP /ʃI:P/	ʃεP	PʌK
28.	SOCK /SɒK/	SʊK	DI:Z
29.	SPOON /SPU:N/	SPɔ:N	TʃεD
30.	TEETH /TI:θ/	Tɪθ	Dʒɔ:L