Cognition 152 (2016) 53–60

Contents lists available at ScienceDirect

Cognition

journal homepage: www.elsevier.com/locate/COGNIT

Original Articles

What infants know about the unsaid: Phonological categorization in the absence of auditory input

Céline Ngon [a], Sharon Peperkamp [a,b,*]

[a] Laboratoire de Sciences Cognitives et Psycholinguistique (Département d'Etudes Cognitives, Ecole Normale Supérieure – PSL Research University/Ecole des Hautes Etudes en Sciences Sociales/Centre National de la Recherche Scientifique), Paris, France
[b] Maternité Port-Royal, Université Paris Descartes, Paris, France

ARTICLE INFO

Article history:
Received 10 October 2013
Revised 14 March 2016
Accepted 19 March 2016

Keywords:
Language development
Phonology
Lexical access
Categorization
Output representations
Inner speech

ABSTRACT

Acquiring a lexicon constitutes an essential step in early language development. From an early age on, infants store words with well-specified phonological representations, and they can spontaneously activate these representations on the basis of visual information only (Mani & Plunkett, 2010a, 2011). To what extent can infants inspect and categorize phonological representations in the absence of auditory input? The present study focuses on words that infants comprehend but do not attempt to pronounce yet, and introduces a novel methodology based on anticipatory eye-movements. In two experiments, 21-month-old French-learning infants were silently presented with images of familiar objects whose labels they comprehended but did not pronounce yet. We tested whether they could activate the phonological representation of these labels and categorize them based on their length. Infants' performance exceeded chance when the target words were mono- and trisyllabic, but not when they were mono- and disyllabic. Thus, even in the absence of auditory input infants can activate the phonological representation of words they do not pronounce yet, and use this representation to perform a categorization based on word length, provided the length difference is substantial.

1. Introduction

Building a lexicon constitutes an essential step in early language acquisition. During the first two years of life, infants develop perceptual skills that allow them to recognize word forms, as well as productive capacities for pronouncing those words. Typically, word comprehension starts earlier than word production: whereas infants show some signs of word comprehension as early as 6–9 months of age (Bergelson & Swingley, 2012; Tincoff & Jusczyk, 1999), they only utter their first words around 12 months (Vihman, 1996).

While adding words to their lexicon, infants develop phonological representations for these words. A host of evidence shows that as early as their second year of life, such representations are well-specified (Bailey & Plunkett, 2002; Mani, Coleman, & Plunkett, 2008; Mani & Plunkett, 2007, 2010b; Skoruppa, Mani, Plunkett, Cabrol, & Peperkamp, 2013; Swingley, 2009; Swingley & Aslin, 2000, 2002; White & Morgan, 2008; Zesiger, Dupuis Lozeron, Lévy, & Frauenfelder, 2011). These studies show that infants are sensitive to mispronunciations in both consonants and vowels: they recognize a target object better when its label is correctly pronounced ('Where's the baby?') than when it is mispronounced ('Where's the vaby?'). Thus, infants perceive phonetic details and store them in their phonological representation of words from an early age. In sharp contrast to these detailed input representations stand infants' approximate and highly variable early word productions. Some of this variability is due to systematic alterations, including sound substitution (for example when a target word consonant takes on features of another consonant, a phenomenon referred to as consonant harmony, e.g. *guck* for *duck*; see e.g. Goad, 1997; Vihman, 1978; Pater & Werle, 2003) and word truncation (when a syllable of a word is omitted, e.g. *nana* for *banana*; see e.g. Allen & Hawkins, 1978; Demuth, 1995; Fikkert, 1994; Gerken, 1994; Ingram, 1978; Pater, 1997; Smith, 1973). It is not until around 6 years of age that most words are pronounced correctly (Sander, 1972; Vihman, 1996).

In models of word production, generating a word begins with the selection of a lemma and the retrieval of the associated word form, including its phonological representation. In adults, the activation of phonological representations is rapid and automatic, and it takes place even without the intention to speak. This follows from research with visual search tasks, in which participants

* Corresponding author at: Laboratoire de Sciences Cognitives et Psycholinguistique, Pavillon Jardin, 29 rue d'Ulm, 75005 Paris, France.
E-mail address: sharon.peperkamp@ens.fr (S. Peperkamp).

http://dx.doi.org/10.1016/j.cognition.2016.03.014
0010-0277/

must identify a target object among a set of objects shown on a screen. Crucially, adults are slower to identify an object in the presence of a distractor whose label is homophonous or phonologically similar to the label of the target object. This phonological interference effect can only be due to the automatic activation of the object labels (Görges, Oppermann, Jescheniak, & Schriefers, 2013; Meyer, Belke, Telling, & Humphreys, 2007). There is evidence that infants who have just started to pronounce words likewise activate phonological representations of unnamed objects (Mani & Plunkett, 2010a, 2011). In these experiments, 18- and 24-month-old English-learning infants heard a label and had to recognize the target object to which it referred (presented side-by-side with a distractor object). Crucially, infants were primed with a silent presentation of an object whose label either started with the same phoneme as the label of the target object or a different one (e.g., target: cat; related prime: cup; unrelated prime: teeth). Infants' recognition was found to be significantly different between related and unrelated trials. That is, 18-month-olds looked longer to the target than to the distractor and were faster to switch from the distractor to the target in related than in unrelated trials; 24-month-old showed the reverse pattern of results, and hence behaved similarly to the adults in the studies mentioned above, whose recognition of a target object was also inhibited by the presence of a phonologically related prime (Görges et al., 2013; Meyer et al., 2007). Thus, infants spontaneously activated the phonological representation of the labels of the silently presented prime objects. It remains to be investigated whether this spontaneous activation depends upon infants' capacity to overtly pronounce the word. Indeed, Mani and Plunkett (2010a, 2011) did not consider infants' expressive vocabulary; given that the primes were familiar monosyllables, it is likely that 24-month-olds already pronounced most of them, and that the 18-month-olds produced at least some.

Infants' capacity to spontaneously activate phonological representations raises a further question: to what extent can infants inspect and categorize these representations? To our knowledge, this question has been addressed only with six- and seven-year old children. Specifically, when shown a set of three pictures, six-year-olds can indicate the picture whose label rhymes with an auditorily presented word, and seven-year-olds can also indicate the picture whose label has the largest number of sounds (Lundberg, Frost, & Petersen, 1988; note that the article does not mention whether the children pronounce the labels before giving a response). In the present study, we examine the capacity to categorize phonological representations of unnamed objects at a much younger age. Specifically, using an implicit anticipatory eye-movement paradigm (McMurray & Aslin, 2004), we investigate whether 21-month-old infants, who do not pronounce many words yet, can categorize the labels of familiar objects that are presented in silence according to whether they are short or long. As categorization requires activation, we also test whether infants can activate phonological representations of unnamed objects even if they do not pronounce their labels overtly. Thus, we use words that according to parental report are known but not yet pronounced.

2. Experiment 1

In this experiment, infants have to categorize monosyllabic and trisyllabic labels. Considering the novelty of the paradigm and the difficulty of the combined activation-categorization task, the 1:3 ratio seems an appropriate starting point. This is the highest possible ratio we can test, since French infants do not know a sufficient number of words with more than three syllables.

Table 1
Mean number of segments and mean durations of mono- and trisyllabic words in Experiment 1. SEs are shown between parentheses.

	Number of segments	Duration
Monosyllable	2.9 (0.1)	738 (46.7)
Trisyllable	6.6 (0.2)	966 (42.0)
Difference	*3.7**** *(0.2)*	*228**** *(44.2)*

*** $p < 0.001$ (t-test).

2.1. Methods

2.1.1. Participants

Thirty-one 21-month-old monolingual French-learning infants from Paris participated (9 boys, mean age = 21;17, range = 20;1–22;1). Parental consent was signed prior to testing. Six additional infants were tested but excluded from analysis due to fussiness before reaching the test phase (5) or experimental error (1).

2.1.2. Stimuli

We selected 30 monosyllabic and 30 trisyllabic words representing unambiguously recognizable objects. Both monosyllabic and trisyllabic lists were matched with regard to semantic category: they contained as many objects corresponding to animals (e.g. *chat* 'cat', *papillon* 'butterfly'), food items (e.g. *oeuf* 'egg', *clémentine* 'clementine'), and artifacts (e.g. *lit* 'bed', *parapluie* 'umbrella'). All words were produced in an infant-directed manner by an adult female native speaker of French. Information about mean number of segments and durations is provided in Table 1. In addition, we selected 60 color drawings depicting the objects.

2.1.3. Procedure

We used an anticipatory eye-movement paradigm, consisting of a learning phase and a test phase (McMurray & Aslin, 2004). Personalized scripts were programmed separately for each infant based on their individual comprehension and production vocabularies.

To obtain vocabulary reports, we asked parents to fill out a questionnaire consisting of the selected word list (see Section 2.1.2) which they had to send back a few days prior to test. For each word parents had to indicate whether they thought their child could comprehend the word, pronounced the word and if so, how it was pronounced. For the test phase, we selected images of objects that infants recognized and whose label they comprehended but did not pronounce, according to parental report.[1] Objects for the learning phase were selected among those remaining in the vocabulary list after assignment of the objects in the test phase; that is, they could be either unknown, known and pronounced, or known but not pronounced by the infant.

The trial design was adapted from Kovács and Mehler (2009). During the learning phase, trials began with the central presentation of an image for 1500 ms, while an isolated auditory label for the depicted object was played simultaneously. The offset of the image was followed by two white squares on each side of the monitor for 1000 ms. Next, the same object reappeared within one of the two squares for 1500 ms, again accompanied by its auditory label. For half of the infants, objects with a monosyllabic label always reappeared on the left side of the screen and objects with a trisyllabic label on the right side; for the other half, it was the

[1] The test stimuli for 17 infants included one or more monosyllabic words they already pronounced. In addition, the test stimuli for 13 infants included one or more trisyllabic words that they pronounced in a truncated, monosyllabic, form (e.g., *clémentine* pronounced as *tine*). All trials with a pronounced mono- or trisyllabic word were omitted from the analyses; this concerned 13.7% of the test trials (9.6% with monosyllabic and 4.2% with trisyllabic words).

reverse. The 1000 ms delay during which the two side-by-side white squares were being presented allowed infants to anticipate the side on which the object would reappear. In order to do this successfully, they had to learn the association between the label's length and the side of the screen associated to that length.

In the following test phase, infants saw *novel* objects. Importantly, their labels were comprehended but not pronounced by the infants. Test trials included the same sequential displays as learning trials, except for two aspects: First, the delay for anticipation was doubled to 2000 ms, in order to give them more time to anticipate. Second, the object was presented **in silence** at the beginning of the trial. Thus, in order to successfully anticipate the side on which the object would reappear, infants had to (1) internally activate the object label themselves and (2) inspect its length. The object's label was played when the object reappeared at the end of the trial, in order to reinforce learning throughout the test and to maintain infants' focus on the task (Fig. 1).

The learning phase always contained 30 different trials, 15 with a monosyllabic and 15 with a trisyllabic word. The test phase also contained equal numbers of trials of both types, but the total number of test trials varied across infants and depended on how many words the infant comprehended but did not pronounce yet (max: 30). In both the learning and the test phase, trials were presented in a pseudo-random order, such that no word from the same type was presented more than three times in a row. The experiment was run using Lincoln Infant Lab Package (Meints & Woodford, 2008).

2.2. Results and discussion

Trials were coded offline frame-by-frame from the offset of the first image to its reappearance. The resulting time window analysis was of 1000 ms in the learning phase and of 2000 ms in the test phase. Infants' gazes were coded as left look, right look, or other look (all coded data for this and the next experiment are available as supplementary material). Data from 13% of the infants were also coded by a second coder (intercoder reliability: 92%, Cohen's Kappa: 0.86). Trials that contained neither a left nor a right look were excluded from the analyses. Two individual accuracy measures were computed based on infants' left and right looks (hence excluding the other looks): initial accuracy, defined as proportion of first fixations to correct side, and overall accuracy, defined as proportion looking time to correct side. Both scores were scaled such that they ran from −1, corresponding to 100% incorrect anticipations, to +1, corresponding to 100% correct anticipations, with 0 representing performance at chance level. To test whether infants' mean accuracy scores were significantly above chance, we ran one-sided t-tests. Given that the number of analyzable test trials varied widely across infants (mean = 14.9; range = 4–27), both because the number of trials was variable and because infants did not necessarily show at least one left or right look in all trials, we used a weighted version. In this version, the contribution of individual infants' accuracy scores to the t-statistic is proportional to their number of anticipated trials.

During the learning phase, infants showed at least one left or right look in 81% of the trials. Infants' mean scores were not significantly above chance for either initial accuracy (mean = −0.02, SE = 0.03, $t(30) < 1$) or overall accuracy (mean = −0.002, SE = 0.03, $t(30) < 1$, ns). During the test phase, they showed at least one left or right look in 83% of the trials. The mean latency of the first look was 625 ms (SD: 141 ms), and its mean duration 819 ms (SD: 238 ms). Infants' mean accuracy scores were significantly above chance, considering both initial accuracy (mean = 0.12, SE = 0.04, $t(30) = 3.25$, $p = 0.0014$; see Fig. 2, left panel) and overall accuracy (mean = 0.06, SE = 0.03, $t(30) = 1.69$, $p = 0.050$).

Thus, infants anticipated significantly more often to the correct side than to the incorrect side, showing that they were able to categorize the labels of silently presented images corresponding to mono-and trisyllabic words according to their length. As categorization requires activation of the words' phonological representation, these results also show that in the absence of auditory input, infants activate phonological representations of familiar words that they do not yet pronounce. Note that although scores for both initial accuracy and overall accuracy are significant, the former are twice as high as the latter. We hypothesize that this is because a 2000 ms time window before the reappearance of the object is long enough for infants to change gaze direction. Thus, given that they do not see the object appear on the side they first orient towards, infants would change their gaze before the end of the time window. Accordingly, first fixations constitute a more reliable measure than proportion looking time for assessing categorization.

There is one caveat: an alternative explanation might be that infants categorized the labels of the two word types based on a frequency rather than a length difference. Using a French corpus of speech directed to infants under 24 months (over 285.000 word tokens; see Ngon et al. (2013) for details), we indeed found that the monosyllabic words we used are significantly more frequent than the trisyllabic ones in infants' speech input (mean number of occurrences for monosyllables: 104, for trisyllables: 31; $t(58) = 4.26$, $p < 0.0001$). Thus, this raises the question of whether infants' relative experience with the objects' labels might have influenced their performance. We therefore reran our analyses on a subset of trials in which the two word types are matched in mean log frequency. To select this subset we proceeded as follows: For each infant, among the attempted test trials we ordered the monosyllabic words from least to most frequent and the trisyllabic words from most to least frequent. We then selected the least frequent monosyllabic word and the most frequent trisyllabic word, and added data points for each word type (so as to obtain equal numbers of monosyllabic and trisyllabic trials). For half of the infants, we stopped selecting trials when the mean log frequency of monosyllables was just below that of the trisyllables, for the other half, when it was just above. Consequently, monosyllabic and trisyllabic trials were matched in log frequency overall. Following this procedure, we obtained a subset containing 66.5% of the original dataset. Three infants did not contribute any datapoints, since for them there was zero overlap between the frequencies of the monosyllables and those of the trisyllables (i.e. the most frequent monosyllable was less frequent than the least frequent trisyllable). As before, we compared mean accuracy scores against chance using a weighted version of the t-test (across infants, the number of analyzable trials varied between 2 and 22, with a mean of 11.0). Even under these more stringent conditions, infants' mean performance significantly exceeded chance for both initial accuracy (mean = 0.17, SE = 0.05, $t(27) = 3.60$, $p = 0.0006$; see Fig. 2, right panel), and overall accuracy (mean = 0.08, SE = 0.04, $t(27) = 2.11$, $p = 0.022$).

Thus, these restricted analyses allow us to reject the alternative interpretation of our data: infants' successful performance is not due to a possible sensitivity to the frequency difference between the two word types, but, rather, is evidence of their sensitivity to the length difference.

In the next experiment, we examine how sensitive infants are to length differences, by testing whether they can categorize mono- vs. disyllables. If their length estimations are noisy, they will make more errors with this smaller ratio, and hence we expect that their performance will be below or equal to that of infants in Experiment 1.

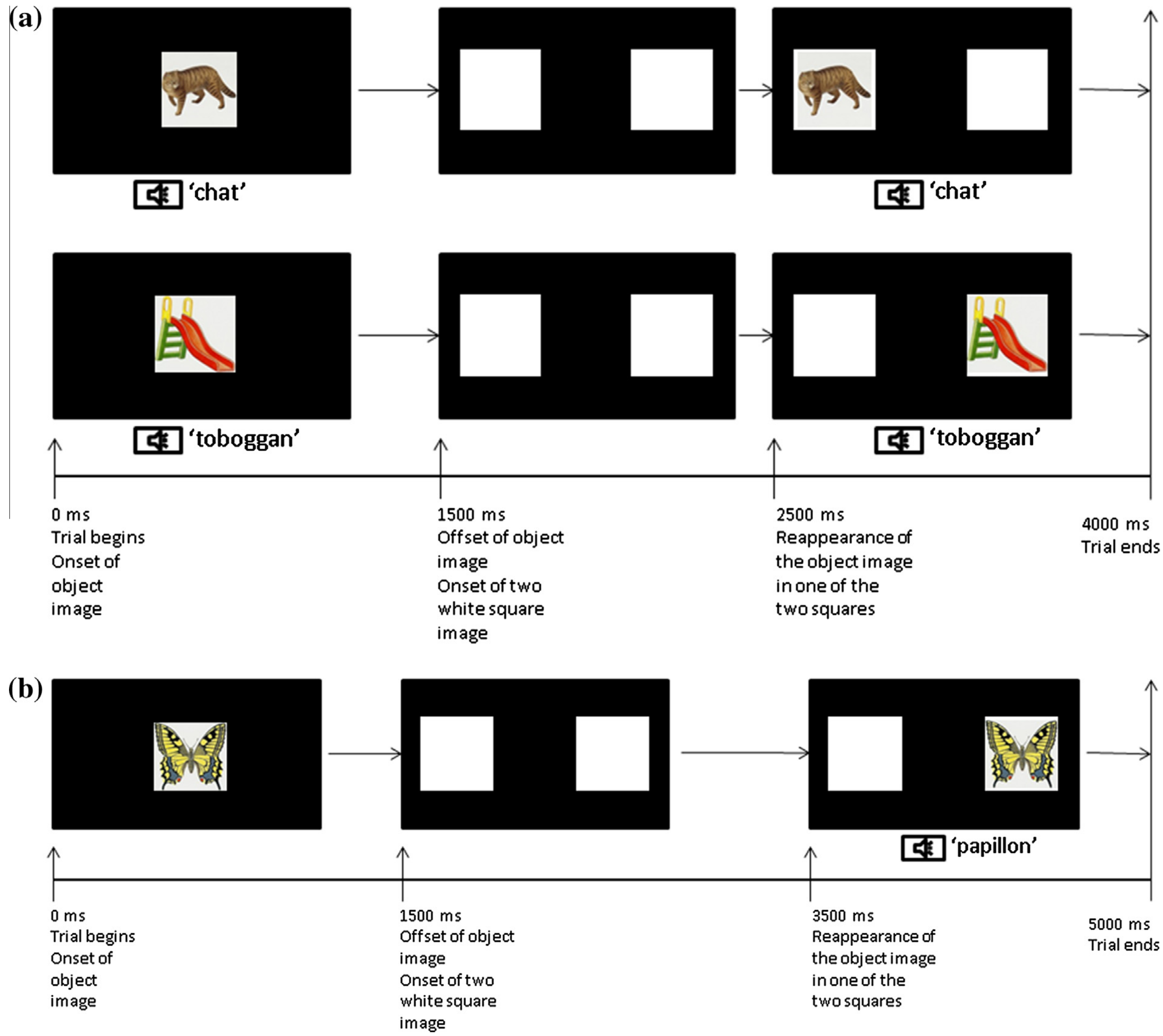

Fig. 1. (a) Time line of two typical learning trials. Each trial included the following displays, in sequence: central presentation of an object (1500 ms), two white squares (1000 ms), and reappearance of the same object in one of the squares (1500 ms). An auditory sound corresponding to the label for the object was played during the first and second presentation of the object. (b) Time line of a typical test trial: central presentation of an object (1500 ms), two white squares (2000 ms), and reappearance of the object in one of the squares (1500 ms). An auditory sound corresponding to the label of the object was played only during the second presentation of the object; the first presentation was silent.

3. Experiment 2

3.1. Methods

3.1.1. Participants

Thirty-one 21-month-old French-learning infants participated[2] (8 boys, mean age = 20;28, range = 20;0–21;26). None of them had participated in Experiment 1. Three additional infants were tested but excluded from analysis due to fussiness before reaching the test phase (2) or experimental error (1).

3.1.2. Stimuli

We selected 44 monosyllabic (e.g. *chat* 'cat', *oeuf* 'egg', *lit* 'bed') and 44 disyllabic words (e.g. *lapin* 'rabbit', *carotte* 'carrot', *voiture* 'car'). Again, both monosyllabic and disyllabic lists were matched with regard to semantic category and all words were produced in an infant-directed fashion by the same speaker as in Experiment 1. Information about mean number of segments and durations is provided in Table 2. In addition, we selected 44 color drawings depicting the objects.

Table 2
Mean number of segments and mean durations of mono- and disyllabic words in Experiment 2. SEs are shown between parentheses.

	Number of segments	Duration
Monosyllable	2.8 (0.1)	707 (25.3)
Disyllable	4.9 (0.1)	917 (29.2)
Difference	*2.1**** *(0.1)*	*210**** *(33.1)*

*** $p < 0.001$ (*t*-test).

3.1.3. Procedure

The experimental design was identical to that in Experiment 1, except that infants had to categorize monosyllabic and disyllabic words and that their parents filled out a questionnaire containing the list of mono- and disyllabic words described above. As in

[2] Five more infants could not be included in the sample, because according to the word list filled out by their parents, their vocabulary contained no words that they understood but did not yet pronounce.

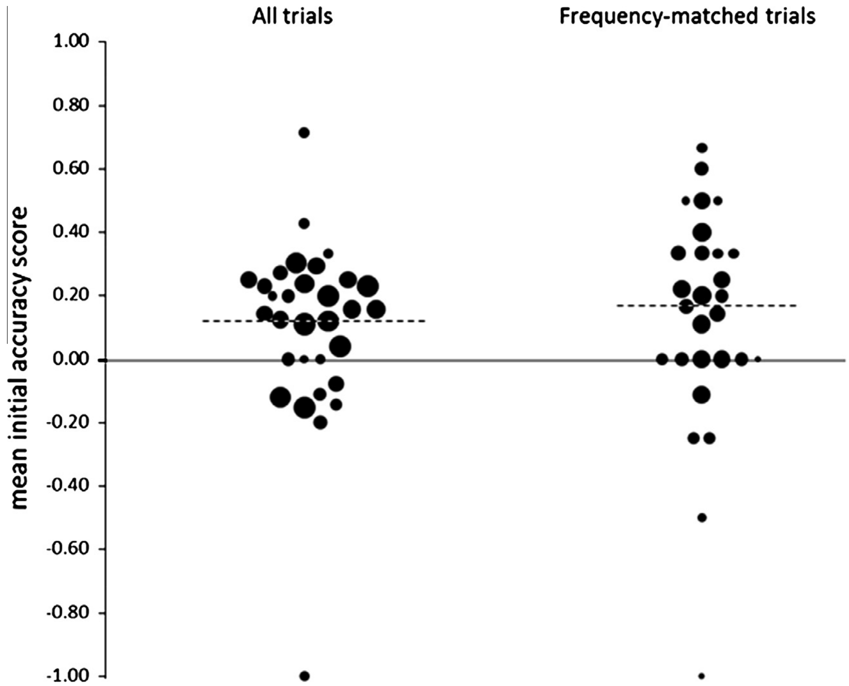

Fig. 2. Individual mean initial accuracy scores in Experiment 1 for all trials (left) and for a subset of the test trials with matched frequencies between monosyllabic and trisyllabic words (right). Individual points have been jittered along the x-axis, with point sizes being proportional to their weight. Dashed lines indicate the weighted group means.

Experiment 1, the test phase contained images of objects infants recognized and whose label they comprehended but did not pronounce according to parental report,[3] while the learning phase included images of objects they did not know, knew and pronounced, or knew but did not pronounce. Since infants tend to know more disyllabic than trisyllabic words, the initial set of words from which training and test items were chosen was larger (see Section 3.1.2); the number of words during the learning phase and the maximum number of words during the test phase, however, were the same as in Experiment 1, i.e. 30. Moreover, care was taken to make the learning phase as similar as possible to the one in Experiment 1. In particular, as the ratios of words that infants do not know, know and pronounce, or know but do not pronounce might influence how well infants learn to categorize short vs. long words, we used on average the same ratios as in Experiment 1.

3.2. Results and discussion

The data from one infant were discarded, since due to a coding error two of the first three test trials (out of 14) contained an object that had been shown on the incorrect side of the screen.

As in Experiment 1, the anticipation time window was coded offline frame-by-frame for each trial. Data from 10% of the infants were also coded by a second coder (intercoder reliability: 93%, Cohen's Kappa: 0.89). The data were analyzed in the same way as those in Experiment 1. Specifically, given the variable number of test trials across infants (mean = 20.0; range = 8–30), infants' performance was compared to chance using a weighted t-test. During the learning phase, infants showed at least one left or right look in 80% of the trials. Their mean scores were not significantly above chance for either initial accuracy (mean = 0.02, SE = 0.03, $t < 1$) or overall accuracy (mean = 0.02, SE = 0.04, $t < 1$). During the test phase, infants showed at least one left or right look in 87% of the trials. The mean latency of the first look was 649 ms (SD: 160 ms), and its mean duration 920 ms (SD: 239 ms). Their mean accuracy scores against chance were significant for neither initial accuracy (mean = 0.03, SE = 0.04, $t < 1$; see Fig. 3) nor overall accuracy (mean = 0.003, SE = 0.03, $t < 1$). Thus, infants failed to categorize mono- vs. disyllables.

It should be noted that there is a significant age difference between infants in Experiment 1 and those in Experiment 2, the infants in Experiment 1 being on average 20 days older than those in the Experiment 2 ($t(59) = 4.80$, $p < 0.0001$). However, age and initial accuracy score did not correlate in Experiment 1 (weighted $r_{\text{exp1}} = -0.075$, $p > 0.1$) and only marginally so in Experiment 2 ($r_{\text{exp2}} = 0.31$, $p = 0.09$). Thus, we tentatively conclude that this age difference cannot explain the difference in performance between the two experiments. Rather, we attribute infants' failure in the present experiment to the fact that a 1:2 ratio is smaller than a 1:3 ratio. The results thus suggests that infants' estimations of length differences are noisy, allowing them to categorize mono- vs. trisyllables but not mono- vs. disyllables.

4. General discussion

It is well known that as early as during the second year of life, infants have detailed phonological representations for familiar words (e.g., for French see Skoruppa et al., 2013; Zesiger et al., 2011). These representations can be activated in the absence of auditory input. Indeed, Mani and Plunkett (2010a, 2011) showed that 18- and 24-month-old infants' word recognition undergoes priming or interference, respectively, from the silent presentation of an object whose label starts with the same phoneme as the label of the target object. In the present article we expanded on this research in two ways. First, while Mani and Plunkett (2010a, 2011) used highly familiar words that the infants probably already pronounced, we examined the activation of words that infants

[3] The test stimuli for 14 infants included one or more disyllabic words that they pronounced in a truncated, monosyllabic, form (e.g., *voiture* pronounced as *ture*). Trials concerning such a truncated bisyllabic word accounted for 5.9% of the test trials; they were omitted from the analyses.

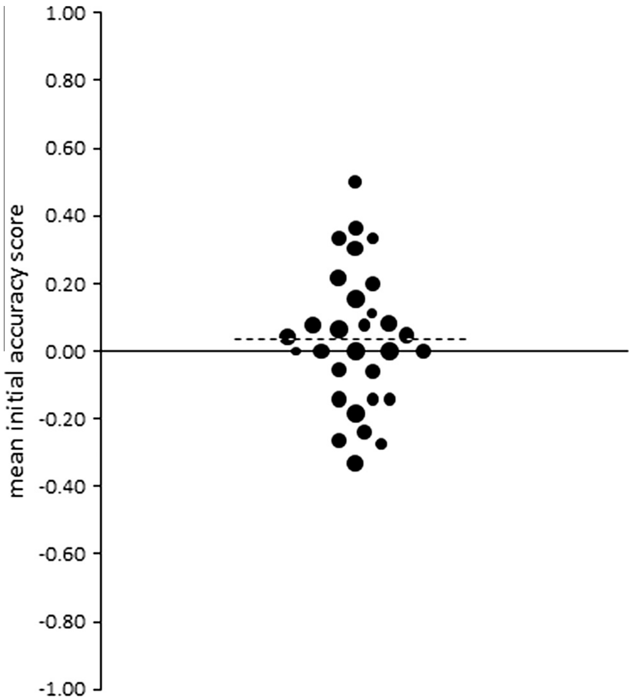

Fig. 3. Individual mean initial accuracy scores for the test phase in Experiment 2. Individual points have been jittered along the x-axis, with point sizes being proportional to their weight. The dashed line indicates the weighted group mean.

comprehend but - according to parental report - do not yet attempt to pronounce. Second, we investigated whether infants can categorize spontaneously activated words according to a phonological criterion. To our knowledge, phonological categorization of words in the absence of auditory input has not been shown in children younger than six years (Lundberg et al., 1988). Here, we tested 21-month-old infants on a combined activation-categorization task, using an adapted version of the anticipatory eye-movement paradigm (McMurray & Aslin, 2004). Infants were presented with pictures of objects whose labels they know but do not yet pronounce. They showed successful categorization of these labels according to whether they are mono- or trisyllabic, but not according to whether they are mono- or disyllabic.

Importantly, the lists of objects with respectively mono- and multisyllabic labels presented during the learning phase were matched with regard to semantic category. Hence, infants could not rely on a semantic categorization to learn the task. Moreover, an analysis of a frequency-matched subset of the trials in the test phase of Experiment 1 allowed us to exclude the possibility that infants performed the categorization task on the basis of frequency rather than length of the objects' labels. These findings, then, provide the first evidence that in the absence of auditory input infants can activate phonological representations for unpronounced words, and, moreover, that they can use these representations to categorize the words according to their length, provided the length difference is substantial (i.e., corresponds to a two-syllable difference).

It should be noted that our conclusion that infants can spontaneously activate phonological representations of words before pronouncing them overtly is only as robust as the parental questionnaire data are trustworthy. That is, for each infant we used test items that according to the parents were not yet pronounced. Even though we used a conservative criterion - items about which parents expressed doubts were not included - infants might have been tested on some items that they already pronounced. Much research has compared parental questionnaire data to measures of infants' expressive language capacity, showing that they are a reliable indicator of the relative size of infants' expressive vocabulary (for infants of around 20 months of age, see Bates, Bretherton, & Snyder, 1988; Corkum & Dunham, 1996; Dale, Bates, Reznick, & Morisset, 1989; Thal, Jackson-Maldonado, & Acosta, 2000), and that, if anything, parents tend to underestimate their infant's expressive vocabulary (Ring & Fenson, 2000). However, evidence that parents can accurately report which words are pronounced and which ones are not yet pronounced is lacking. Our questionnaires were relatively short; for most parents they only contained the items preselected for use in the experiment (60 for infants in Experiment 1 and 88 for infants in Experiment 2). This might have induced parents to respond more carefully than if they had been given a complete language assessment questionnaire, which typically contains up to 700 items. Still, we cannot be certain that they could reliably indicate which words their infant did not yet pronounce.

In Experiment 1, what type of length information did infants rely on to learn the categorization of monosyllabic vs. trisyllabic words during the learning phase? One possibility is of course that they were sensitive to the difference in number of syllables. Alternatively, they might have relied on the difference in number of segments, as the trisyllabic words contained more segments than the monosyllabic ones. A third possibility is that infants relied on a low-level cue, i.e. the difference in raw acoustic duration; on average, the trisyllabic words indeed also had a longer duration than the monosyllabic ones. These possibilities are not mutually exclusive: infants might have learned the categorization based on multiple types of information available in the stimuli. Infants' failure with mono- vs. disyllabic words in Experiment 2 provides some more insight into this question: If infants are sensitive to the number of syllables, categorizing mono- vs. disyllables is likely more difficult than categorizing mono- vs. trisyllables because a ratio of 1:2 is smaller than a ratio of 1:3.[4] The same holds if infants are sensitive to the number of segments. Indeed, considering the stimuli we used, we find that the difference in number of segments between mono- and trisyllables is significantly greater than that between mono- and disyllables (3.7 vs. 2.1, $F(2,72) = 30.98$, $p < 0.0001$). By contrast, contrary to what one might expect, the durational difference between our mono- and trisyllabic stimuli was **not** significantly greater than that between our mono- and disyllabic stimuli (228 vs. 210 ms, $F(2,72) < 1$). Thus, while from a phonological point of view the learning phase was more difficult in Experiment 2 than in Experiment 1, from a durational point of view there was no difference. This, then, suggests that infants relied on phonological length - i.e. number of syllables or segments - and not (or not only) on acoustic duration.[5]

One caveat is in order, though. In Experiment 2, infants might have had more difficulty either to discover the correct association during the learning phase and/or to categorize the words they activated during the test phase. Thus, we cannot exclude the

[4] Previous work investigating infants' representation of numbers for objects with different experimental paradigms (e.g. looking time, manual search, or locomotor approach) showed that 10–14 months infants can discriminate between arrays of 1 vs. 2 objects as well as of 1 vs. 3 objects (Feigenson & Carey, 2005; Feigenson, Dehaene, & Spelke, 2004). Categorizing words according to their number of syllables, however, is certainly more challenging than categorizing arrays according to their number of discrete objects, and it seems likely that a 1:2 ratio is more difficult than a 1:3 ratio when it comes to number of syllables in words. In particular, if infants' estimation of number of syllables is noisy, they will make more errors with a 1:2 than with a 1:3 ratio.

[5] To the extent that number of syllables, not segments, defines the metrical structural of words, it seems more plausible that infants relied on number of syllables than on number of segments. Indeed, metrical structure is a perceptually salient feature of speech, to which even newborns are sensitive: they discriminate lists of di- and trisyllabic items but fail to discriminate lists of disyllabic items that are either four or six segments long (Bijeljac-Babic, Bertoncini, & Mehler, 1983).

possibility that they had inferred the correct categorization by the end of the learning phase but failed the task during the test phase. In this case, the possibility that during the learning phase they relied only on acoustic duration remains open. In order to shed light on this issue, it would be interesting to lengthen this phase (and even to dispense with the test phase altogether) in future research. Recall that neither of our experiments showed an effect of learning during the learning phase itself. With a longer learning phase, we expect to see above-chance anticipations for mono- vs. trisyllables towards the end of this phase; for mono- vs. disyllables, we have no prediction other than that performance should not be better than that on mono- vs. trisyllables. Overall, the results would inform us as to whether infants are able to learn to categorize auditorily presented mono- vs. disyllables, and if so, whether there is a difference in learning difficulty compared to mono- vs. trisyllables.

As to the silent test phase, it is impossible to infer with certainty which type of cue infants relied on to categorize the object labels. If they learned the side of screen assignments based on phonological length during the learning phase, we expect a priori that they used the same cue but now applied to phonological representations they spontaneously activated. To the extent that phonological representations might include information on raw duration, we cannot exclude, though, that they relied on such information. Indeed, while in our stimuli the sets of disyllabic and trisyllabic words were not significantly different in duration, infants acquire their phonological representations based on the speech they hear in their daily lives, in which di- and trisyllabic words most likely do have different durations. The precise nature of the cues infants rely on in the absence of auditory input remains a question for future research. Regardless, the present results provide novel insight into the developing lexicon, as we show for the first time that infants are able to spontaneously activate and categorize phonological representations for words that they have not yet begun to pronounce.

A further question is *how* infants' activation of phonological representations of unnamed object labels comes about. Much of the adult speech production literature considers this type of activation to be an implicit, automatic step during lexical access (Morsella & Miozzo, 2002; Navarrete & Costa, 2005; Meyer & Damian, 2007). Alternatively, Meyer et al. (2007) raise the possibility that adults actively generate the labels of unnamed objects in inner speech. In a similar vein, Mani and Plunkett (2010a) consider their results with the picture-based phonological priming paradigm to be evidence for implicit naming in infants, and Perrone-Bertolotti, Rapin, Lachaux, Baciu, and Loevenbruck (2014), in a review article on inner speech, consider both the present findings and those by Mani and Plunkett (2010a) as evidence for inner speech in infants. They point out, moreover, that this capacity could facilitate infants' oral language development. As we used words that infants do not yet pronounce, the inner speech interpretation would indeed imply that infants have output phonological representations of words before they start pronouncing them. This would be in line with experimental evidence that young children can internally generate the correct phonological form of words they pronounce incorrectly: Brett, Chiat, and Pilcher (1988) found that 5-year-old children with phonological disorder who replace word-initial /k/ by /t/ (e.g., producing *tap* instead of *cap*), succeed at distinguishing the contrast between /k/ and /t/ in a picture categorization task, in which they were explicitly asked to "think" of the labels representing the images so as to internally inspect their initial consonant. The authors concluded that children have correct output phonological representations of these mispronounced words and that pronunciation errors (velar fronting in their case) occur during the following stage of articulatory planning. Of course, in our case infants' output phonological representations would not necessarily be fully correct, but even if they were rudimentary their development would be more advanced than suggested by the absence of covert articulation.

The present study opens the way to use our innovative paradigm for further explorations of the development of phonological representations in infants and young children. Indeed, the combined activation-categorization task is quite demanding, but despite its high cognitive load it is very engaging for infants, as shown by our low attrition rates (on average 10%). One question that could be examined using the same methodology concerns the featural structure of segments. For instance, infants could be trained to categorize words starting with a subset of stops (e.g., /p,b,k,g/) vs. those starting with a subset of fricatives (e.g. /f,v,ʃ,ʒ/), and tested on their categorization of words starting with stops and fricatives not presented during the learning phase (/t,d,s,z/). If infants represent subsegmental structure, they should generalize and hence categorize the test words as stop-initial (/t,d/) vs. fricative-initial (/s,z/).

Under the assumption that infants generate phonological output representations by means of inner speech, another particularly interesting question is that of truncated words in early production patterns. Some French-learning infants go through a stage in which they pronounce only the final, stressed, syllable of certain multisyllabic words (e.g., *gan* for *toboggan*, 'slide', Demuth & Johnson, 2003). Similar truncations have been documented in the speech of infants learning a variety of languages, and they have generally been interpreted as evidence for inaccurate phonological representations (e.g., Demuth, 1995; Fikkert, 1994; Pater, 1997; Smith, 1973; but see Smith, 2010). By contrast, based on the present findings on words that are not pronounced at all, we expect that infants activate an output representation with adult-like length for truncated trisyllabic words. Thus, they should categorize *toboggan* as long, regardless of whether they pronounce it as a monosyllable or do not attempt to pronounce it at all.[6] Such outcome would be in line with acoustic evidence showing that omitted syllables can leave unperceivable prosodic traces in the speech of infants and young children (e.g. Carter & Gerken, 2004).

To conclude, we introduced a novel experimental paradigm based on anticipatory eye movements that allows us to examine the activation and categorization of phonological representations in the absence of auditory input. We used this paradigm to show that 21-month-old infants can activate the phonological representation of mono- and trisyllabic words they do not yet pronounce, and categorize them according to their length. While the precise nature of spontaneously activated phonological representations as well as the activation mechanism itself remain topics for future research, the present results reveal sophisticated phonological capacities in infants long before they pronounce words in an adult-like fashion.

Acknowledgements

This study was funded by the Agence Nationale de la Recherche (ANR-13-APPR-0012, ANR-10-LABX-0087 IEC, ANR-10-IDEX-0001-02 PSL*). We would like to thank Anne Christophe for stimuli recording, Michel Dutat for technical assistance, Alexandre Cremers for statistical advice, and Marieke van Heugten for comments and discussion.

[6] Recall that some infants in our study were presented with test trials – excluded from the analyses – containing an image of an object they pronounced in a truncated form (see Footnotes 1 and 3). Due to the low number of such trials it is impossible to analyze them separately. Including them for the purposes of the overall analyses, however, does not change the results.

Appendix A. Supplementary material

Supplementary data associated with this article can be found, in the online version, at http://dx.doi.org/10.1016/j.cognition.2016.03.014.

References

Allen, G., & Hawkins, S. (1978). The development of phonological rhythm. In A. Bell & J. B. Hooper (Eds.), *Syllables and segments* (pp. 173–185). Amsterdam: North Holland.

Bailey, T., & Plunkett, K. (2002). Phonological specificity in early words. *Cognitive Development, 17*, 1267–1284.

Bates, E., Bretherton, I., & Snyder, L. (1988). *From first words to grammar*.Cambridge, UK: Cambridge University Press.

Bergelson, E., & Swingley, D. (2012). At 6 to 9 months, human infants know the meanings of many common nouns. *Proceedings of the National Academy of Sciences, 109*, 3253–3258.

Bijeljac-Babic, R., Bertoncini, J., & Mehler, J. (1983). How do 4-day-old infants categorize multisyllabic utterances? *Developmental Psychology, 29*, 711–721.

Brett, L., Chiat, S., & Pilcher, C. (1988). Stages and units in output processing: Some evidence from voicing and fronting processes in children. *Language and Cognitive Processes, 2*, 165–177.

Carter, A., & Gerken, L. A. (2004). Do children's omissions leave traces? *Journal of Child Language, 31*, 561–586.

Corkum, V., & Dunham, P. (1996). The communicative development inventory-WORDS short form as an index of language production. *Journal of Child Language, 23*, 515–528.

Dale, P., Bates, E., Reznick, J., & Morisset, C. (1989). The validity of a parent report instrument of child language at twenty months. *Journal of Child Language, 16*, 239–250.

Demuth, K. (1995). Markedness and the development of prosodic structure. In J. Beckman (Ed.). *Proceedings of the north east linguistic society* (Vol. 25, pp. 13–25). Amherst, MA: GLSA, Univ. of Mass.

Demuth, K., & Johnson, M. (2003). Truncation to subminimal words in early French. *Canadian Journal of Linguistics, 48*, 211–241.

Feigenson, L., & Carey, S. (2005). On the limits of infants' quantification of small object arrays. *Cognition, 97*, 295–313.

Feigenson, L., Dehaene, S., & Spelke, E. (2004). Core systems of number. *Trends in Cognitive Sciences, 8*, 307–314.

Fikkert, P. (1994). On the acquisition of prosodic structure. *HIL dissertations in Linguistics 6*. The Hague: Holland Academic Graphics.

Gerken, L. A. (1994). A metrical template account of children's weak syllable omissions from multisyllabic words. *Journal of Child Language, 21*, 565–584.

Goad, H. (1997). Consonant harmony in child language: An optimality-theoretic account. In S. J. Hannahs & M. Young-Scholten (Eds.), *Focus on phonological acquisition* (pp. 113–142). Amsterdam: John Benjamins.

Görges, F., Oppermann, F., Jescheniak, J., & Schriefers, H. (2013). Activation of phonological competitors in visual search. *Acta Psychologica, 143*, 168–175.

Ingram, D. (1978). The role of the syllable in phonological development. In A. Bell & J. Hooper (Eds.), *Syllables and segments* (pp. 143–155). New York: Elsevier-North Holland.

Kovács, Á., & Mehler, J. (2009). Cognitive gains in 7-month-old bilingual infants. *Proceedings of the National Academy of Sciences, 106*, 6556–6560.

Lundberg, I., Frost, J., & Petersen, O. (1988). Effects of an extensive program for stimulating phonological awareness in preschool children. *Reading Research Quarterly, 23*, 263–285.

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

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

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

Mani, N., & Plunkett, K. (2010b). Twelve-month-olds know their cups from their keps and tups. *Infancy, 15*, 445–470.

Mani, N., & Plunkett, K. (2011). Phonological priming and cohort effects in toddlers. *Cognition, 121*, 196–206.

McMurray, B., & Aslin, R. (2004). Anticipatory eye movements reveal infants' auditory and visual categories. *Infancy, 6*, 203–229.

Meints, K., & Woodford, A. (2008). Lincoln Infant Lab Package 1.0: A new programme package for IPL, preferential listening, habituation and eyetracking Available at:<www.lincoln.ac.uk/psychology/babylab.htm> [WWW document: Computer software & manual].

Meyer, A., Belke, E., Telling, A., & Humphreys, G. (2007). Early activation of object names in visual search. *Psychonomic Bulletin & Review, 14*, 710–716.

Meyer, A., & Damian, M. (2007). Activation of distractor names in the picture-picture interference paradigm. *Memory & Cognition, 35*, 494–503.

Morsella, E., & Miozzo, M. (2002). Evidence for a cascade model of lexical access in speech production. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 28*, 555–563.

Navarrete, E., & Costa, A. (2005). Phonological activation of ignored pictures: Further evidence for a cascade model of lexical access. *Journal of Memory and Language, 53*, 359–377.

Ngon, C., Martin, A., Dupoux, E., Cabrol, D., Dutat, M., & Peperkamp, S. (2013). (Non)words, (non)words, (non)words: evidence for a protolexicon during the first year of life. *Developmental Science, 16*, 24–34.

Pater, J. (1997). Minimal violation and phonological development. *Language Acquisition, 6*, 201–253.

Pater, J., & Werle, A. (2003). Direction of assimilation in child consonant harmony. *Canadian Journal of Linguistics, 48*, 385–408.

Perrone-Bertolotti, M., Rapin, L., Lachaux, J.-P., Baciu, M., & Loevenbruck, H. (2014). What is that little voice inside my head? Inner speech phenomenology, its role in cognitive performance, and its relation to self-monitoring. *Behavioural Brain Research, 261*, 220–239.

Ring, E., & Fenson, L. (2000). The correspondence between parent report and child performance for receptive and expressive vocabulary beyond infancy. *First Language, 20*, 141–159.

Sander, E. K. (1972). When are speech sounds learned? *Journal of Speech and Hearing Disorders, 37*, 55–63.

Skoruppa, K., Mani, N., Plunkett, K., Cabrol, D., & Peperkamp, S. (2013). Early word recognition in sentence context: French and English 2-year-olds' sensitivity to sentence-medial mispronunciations and assimilations. *Infancy, 18*, 1007–1029.

Smith, N. (1973). *The acquisition of phonology: A case study*.Cambridge, MA: Cambridge University Press.

Smith, N. (2010). *Acquiring phonology: A cross-generational case-study*.Cambridge, UK: Cambridge University Press.

Swingley, D. (2009). Onsets and codas in 1.5-year-olds' word recognition. *Journal of Memory and Language, 60*, 252–269.

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

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

Thal, D., Jackson-Maldonado, D., & Acosta, D. (2000). Validity of a parent-report measure of vocabulary and grammar for Spanish-speaking toddlers. *Journal of Speech, Language, and Hearing Research, 43*, 1087–1100.

Tincoff, R., & Jusczyk, P. (1999). Some beginnings of word comprehension in 6-month-olds. *Psychological Science, 10*, 172–175.

Vihman, M. (1996). *Phonological development. The origins of language in the child*. Oxford, UK: Blackwell.

Vihman, M. M. (1978). Consonant harmony: Its scope and function in child language. In J. H. Greenberg (Ed.), *Universals of human language. Phonology* (Vol. 2, pp. 281–334). Stanford: Stanford University Press.

White, K., & Morgan, J. (2008). Sub-segmental detail in early lexical representations. *Journal of Memory and Language, 59*, 114–132.

Zesiger, P., Dupuis Lozeron, E., Lévy, A., & Frauenfelder, U. (2011). Phonological specificity in 12- and 17-month-old French-speaking infants. *Infancy, 17*, 1–19.