Sound symbolism in infancy: Evidence for sound–shape cross-modal correspondences in 4-month-olds

Ozge Ozturk *,1, Madelaine Krehm *,1, Athena Vouloumanos

Department of Psychology, New York University, New York, NY 10003, USA

ABSTRACT

Perceptual experiences in one modality are often dependent on activity from other sensory modalities. These cross-modal correspondences are also evident in language. Adults and toddlers spontaneously and consistently map particular words (e.g., ‘kiki’) to particular shapes (e.g., angular shapes). However, the origins of these systematic mappings are unknown. Because adults and toddlers have had significant experience with the language mappings that exist in their environment, it is unclear whether the pairings are the result of language exposure or the product of an initial proclivity. We examined whether 4-month-old infants make the same sound–shape mappings as adults and toddlers. Four-month-olds consistently distinguished between congruent and incongruent sound–shape mappings in a looking time task (Experiment 1). Furthermore, mapping was based on the combination of consonants and vowels in the words given that neither consonants (Experiment 2) nor vowels (Experiment 3) alone sufficed for mapping. Finally, we confirmed that adults also made systematic sound–shape mappings (Experiment 4); however, for adults, vowels or consonants alone sufficed. These results suggest that some sound–shape mappings precede language learning, and may in fact aid in language learning by establishing a basis for matching labels to referents and narrowing the hypothesis space for young infants.

© 2012 Elsevier Inc. All rights reserved.
Introduction

The mapping between labels and referents has been considered arbitrary in mainstream linguistics (De Saussure, 1916/1983). However, research findings suggest some systematicity between specific labels and referents; adults and toddlers spontaneously and consistently map particular shapes to particular words even when there appears to be no obvious physical basis for the mapping (e.g., Köhler, 1947; Maurer, Pathman, & Mondloch, 2006). However, because adults and toddlers already have significant experience with language mappings in their environment, it is unclear whether these biases are the result of an acquired sensitivity or an initial proclivity. Here we investigated the origins of these sound–shape mapping biases by examining whether 4-month-old infants share the same sound–shape mapping biases as adults and toddlers.

A fundamental tenet of modern linguistics is that there is no systematic relationship between linguistic labels and the meaning they convey (De Saussure, 1916/1983). Some have even proposed that this arbitrary connection between linguistic labels and their referents is a fundamental feature of human language (e.g., Monaghan, Fitneva, & Christiansen, 2011; Ramachandran & Hubbard, 2001) and the basis for its referential power (e.g., Gasser, 2004). Moreover, if the sound structure of words was related in a systematic way to their referents, we would expect similarities across languages, with similar inventories of sounds corresponding to similar types of meaning. This is clearly not the case because even a simple concept such as tree is realized with phonologically distinct words in different languages—ağaç in Turkish, zuhaitza in Basque, árbol in Spanish, and dendro in Greek. Indeed, most linguistic labels have different sound structures in different languages; these labels are conventions agreed on by the speakers of the languages with no apparent systematic relationship to their referents (Hockett, 1977).

However, not all label–referent mappings appear to be arbitrary. Cross-linguistic observations suggest that there are words in natural language whose sounds are systematically related to their meanings. Ideophones—words that are used by speakers to evoke vivid associations with particular sensory perceptions (e.g., smell, color, shape, sound, action, or movement across languages)—are widely attested in the languages of the world (Voeltz & Kilian-Hatz, 2001). West African, East Asian, and South-East Asian languages, and to a lesser extent Amerindian languages, are known for their large inventories of ideophonic vocabulary, but many other languages also make use of ideophones (e.g., Turkish citir citir ‘crispy’; Basque mara mara ‘falling softly, said of rain’; Ewe gbazzaa ‘flat, spreading out over a wide area’; English bling bling ‘glitter, sparkle’). However, the sounds of these words evoke or imitate the meaning, creating systematic nonarbitrary connections.

People also systematically map particular speech sounds to properties of objects cross-linguistically. For instance, across languages, there are phonetic classes of speech sounds that are systematically found in vocabulary related to size (e.g., /i/ vs. /a/ for size as in Ewe kitikitsi ‘small’ vs. ghagbagba ‘big’, Greek micros ‘small’ vs. macros ‘large’) or distance (e.g., words for here are more likely to include an /i/ sound and words for there are more likely to include an /a/ sound; see Tanz, 1971). Experimental findings also support this idea of sound–size relations. Syllables containing low back vowels (e.g., mal) are consistently matched to large objects, whereas syllables containing high front vowels (e.g., mil) are consistently matched to small objects (Sapir, 1929). Recently, cross-modal mapping of sound and size has also been demonstrated in infants, with 4-month-olds matching [o] or [a] to large objects and [i] or [e] to small objects (Peña, Mehler, & Nespor, 2011). Thus, systematic mapping of sound to size is widespread and observed even during infancy.

Cross-modal correspondences between sound and size have been noted for centuries (Descartes, 1641/1986; Gibson, 1966; Ohala, 1997; Walker et al., 2010) and appear to be grounded in physical reality. High front vowels such as [i] and [e] tend to have higher fundamental frequencies ($F_0$) than low back vowels such as [o] and [a] (Whalen & Levitt, 1995). At the same time, objects that are physically thinner or smaller tend to produce a higher pitch than wider larger objects. For instance, the pitch of a cello is lower compared with the pitch made by its smaller cousin, the violin. Similarly, the vocal folds of a larger animal are longer, and longer vocal folds tend to generate sounds that are lower in pitch (Shayan, Ozturk, & Sicoli, 2011; Zbikowski, 1998). Peña and colleagues (2011) proposed that infants may map vowels to object size based on their experience in seeing mouths open to
different extents when humans produce different vowels. The systematic mapping between vowels and object size appears to be grounded on observable physical relationships between the pitch of vowels and the size of objects and, therefore, may simply reflect associations based on participants' prior multisensory experiences.

Another type of cross-modal mapping that has been observed in adults, and recently in infants as young as 3 or 4 months, is a correspondence between auditory pitch and visuospatial height or visual sharpness (Marks, 1987; Walker et al., 2010). Infants (and adults) map higher pitched tones with sharper objects and objects that appear higher in space. Walker and colleagues (2010) proposed that the mapping between sharpness and pitch might be mediated by the hardness of natural materials. Sharp and pointed objects tend to be formed from hard materials, and they tend to produce high-pitched sounds when struck. These natural correspondences are also grounded on observable physical relationships and may reflect prior multisensory experience.

However, there are other cross-modal mappings between linguistic labels and their referents that may rely less on prior multisensory experience in a phenomenon known as sound symbolism. Adult Spanish speakers consistently matched curvy shapes with the novel word ‘baluba’ and angular shapes with another novel word ‘takete’ (Köhler, 1947). These mapping biases were later documented in American college undergraduates and Tamil speakers in India, with the vast majority of both groups (as many as 98%) selecting a curvy shape as ‘bouba’ and an angular shape as ‘kiki’, and this effect became known as the “bouba/kiki effect” (Köhler, 1947). These mapping biases also affect processing; adults are more accurate and faster at classifying a novel object if the category name matches the shape of the object (Kovic, Plunkett, & Westermann, 2010). Electrophysiological responses also differentiate between sound symbolic and non-sound symbolic categories, producing an early neural activation similar to that elicited by highly learned categories, possible evidence of auditory–visual integration (Kovic et al., 2010). Curvy and angular shapes may be related to the shape of a speaker’s lips when producing the vowels (open and rounded or wide and narrowed; see Ramachandran & Hubbard, 2001). However, this would require attending to and mapping two sets of visual properties of two different entities (the object shape and the speaker’s mouth) that are not spatially colocalized or even always temporally copresent—a much less straightforward mapping than noticing natural physical relationships between sounds and objects. Thus, sound–shape correspondences are less easily explained by observable physical relationships but are systematic nevertheless.

Nonarbitrary mappings between sound and shape extends into childhood. When asked to label Köhler-type curvy and angular drawings with the word ‘takete’ or ‘uloomo’, both English-speaking children (11–14 years old) from England and Swahili-speaking children (8–14 years old) from central Africa matched the curvy shape with the word ‘uloomo’ and the angular shape with the word ‘takete’ (Davis, 1961). This systematic sound–shape mapping has even been demonstrated in 2½-year-olds. Children systematically matched ‘bouba’ to curvier shapes and ‘kiki’ to more angular shapes (Maurer et al., 2006).

Both adults and children make systematic cross-modal mappings between linguistic labels and their referents. However, the origins of these sound–shape mappings are unknown. Because adults and toddlers have had significant experience with the statistics of label mappings that exist in their environment, it is unclear whether these systematic pairings are the result of language exposure or the product of an initial proclivity. Young infants who have yet to develop a lexicon have not been exposed to the statistical regularities in their language long enough to be biased by vocabulary-wide systematicity in labels for specific sound–shape patterns; for example, corpus analyses showed that velars are found more in words for angular — rather than rounded — objects. (Monaghan, Mattock, & Walker, 2012). Therefore, if infants share the same sound–shape mappings as adults and toddlers, these biases may precede and perhaps aid in word or category learning, narrowing the massive number of possible sound–shape correspondences that learners face (Imai, Kita, Nagumo, & Okada, 2008; Monaghan et al., 2012; Nygaard, Cook, & Namy, 2009; Parault, 2006; Parault & Schwanenflugel, 2006).

To investigate the origins of these biases, we examined whether prelinguistic infants exhibit the same sound–shape mapping biases as adults and toddlers. Furthermore, we systematically explored whether infant mappings are based on a combination of vowels and consonants, vowels alone, or consonants alone, and we compared these results with those for adults. In Experiment 1, we presented infants with a large red shape that was either curvy or angular, accompanied by a nonsense word.
either ‘bubu’ or ‘kiki’. We predicted that if 4-month-olds who have less language exposure and few, if any, word–object mappings in their receptive vocabulary (Fenson et al., 1994; see also Bergelson & Swingley, 2012) have the same biases as adults and toddlers, they would look differently at trials considered as congruent by adults (pairing the angular shape with ‘kiki’ or the curvy shape with ‘bubu’) and at incongruent trials (pairing the curvy shape with ‘kiki’ or the angular shape with ‘bubu’). If, on the other hand, the biases demonstrated by adults and toddlers are developed through significant language experience, 4-month-olds should look equally at incongruent and congruent trials. In Experiments 2 and 3, we examined whether mappings are based on vowels or consonants. Finally, in Experiment 4, we tested adults on these materials and examined their use of vowels or consonants in making systematic mappings.

Experiment 1

Experiment 1 examined whether prelinguistic infants make the same mappings between nonsense words and shapes as adults (Köhler, 1947; Ramachandran & Hubbard, 2001) and toddlers (Maurer et al., 2006). We presented 4-month-old infants with two sounds: ‘kiki’ and ‘bubu’. These two sounds were paired with a curvy shape or an angular shape. We examined whether infants looked differently at pairings that adults and toddlers in prior studies had rated as incongruent (the curvy shape with ‘kiki’ and the angular shape with ‘bubu’) than at pairings that had been rated as congruent (the angular shape with ‘kiki’ and the curvy shape with ‘bubu’).

Method

Participants

The participants were 12 full-term 4-month-olds (mean age = 4 months 4 days, range = 3 months 22 days to 4 months 20 days, 5 girls and 7 boys). Infants were recruited at birth from local birthing hospitals. A research assistant subsequently contacted the families to participate in the study. Parents were compensated for transportation, and infants were given a small gift for participating. Infants’ parents filled out a demographic information form and reported a range of educational backgrounds; of the 24 parents, 1 had not completed any high school, 1 had completed some high school, 2 had completed some college, 8 had completed college, 6 had completed graduate degrees, and 6 declined to answer. Of the 12 infants, 6 were reported as being White, 4 were of mixed ethnicity, 1 was Hispanic, and 1 parent declined to answer. An additional 13 infants were tested but excluded from the analysis because of fussiness (6), inattentiveness (2), equipment failure (3),2 or experimenter error (2). All procedures were approved by the New York University institutional review board.

Materials

Auditory stimuli. A female English–Turkish bilingual speaker recorded several productions of ‘kiki’ and ‘bubu’ in a sound attenuated room with a Shure SM58 microphone. We decided to use ‘bubu’ instead of ‘buba’, which had been used in previous studies (Maurer et al., 2006), to match the reduplicative structure of ‘kiki’. Three ‘kiki’ and three ‘bubu’ tokens were selected and matched on duration (‘kiki’: $M = 415$ ms, range = $400–425$; ‘bubu’: $M = 408$ ms, range = $400–417$) and modified using Praat (Boersma & Weenink, 2010) to match on amplitude (all tokens were 70 dB) and pitch (‘kiki’: $M = 230$ Hz, range = $227–233$; ‘bubu’: $M = 230$ Hz, range = $228–232$). For each word, the three selected tokens were combined in a semirandomized order, with an interstimulus interval jittered around 1250 ms (ranging from 1000 to 1500 ms), creating a 40-s ‘kiki’ string and a 40-s ‘bubu’ string, each of which contained 24 tokens.

2 The relatively higher attrition rate for Experiment 1 was due to data loss from equipment failure. Specifically, there was no sound played during the experiment. This was corrected in subsequent experiments.

Please cite this article in press as: Ozturk, O., et al. Sound symbolism in infancy: Evidence for sound–shape cross-modal correspondences in 4-month-olds. Journal of Experimental Child Psychology (2012), http://dx.doi.org/10.1016/j.jecp.2012.05.004
Visual stimuli. We created two visual stimuli: one of a curvy shape and one of an angular shape (see Fig. 1). Both shapes were red and presented on a black background. The curvy and angular shapes filled 22.6% and 22.3% of the screen, respectively.

Procedure

Infants were tested in a sound attenuated room. They sat on a parent’s lap 1 m away from a 30-inch Apple monitor. Sounds were played from an M-Audio BX5a speaker located behind the monitor. A Sony DCR-HC96 camera mounted below the monitor recorded the infants and allowed the experimenter to watch the silent video feed in a nearby room. Parents wore Extreme Isolation Ex-25 noise-canceling headphones that played music mixed with spoken words to mask experimental sounds.

Each infant was tested in an infant-controlled cross-modal matching paradigm with sequential presentation of congruent and incongruent trials. The infant controlled both the onset and the offset of the trials. When the infant oriented to the screen, the experimenter started the trial. The trial ended when the infant looked away from the screen for a minimum of 2 s or after a maximum trial length of 40 s. The experiment began with a red flashing light on a black background. The first trial was a pretest trial to familiarize the infant with the procedure. During the pretest trial, the infant saw a shape that had a combination of curvy and angular edges and also heard music. After the pretest trial, the experiment began. The experiment consisted of four identical blocks. Each block consisted of four trials with each of the four possible word/shape pairings: ‘kiki’ with a curvy shape, ‘kiki’ with an angular shape, ‘bubu’ with a curvy shape, and ‘bubu’ with an angular shape. The trials were in semirandom order such

---

Fig. 1. Curvy and angular shapes.
that each pair of stimuli within a block consisted of one congruent trial and one incongruent trial. Whether the first trial was congruent or incongruent was counterbalanced across infants. The red flashing light was presented between trials to draw the infant’s attention back to the screen. An animated video clip was played between the blocks to retain the infant’s attention. The experiment was presented using Habit X software (Cohen, Atkinson, & Chaput, 2004) on an iMac Apple computer, with the experimenter recording online looking time with a key press. Infant looking times were coded offline by a coder naive to the experimental conditions, and the offline data were used in the analyses. Online and offline coding had a correlation of \( r = .78 \) across all three infant experiments, and 25% of trials for each infant were reliability coded by a second offline coder. Data from the two offline coders were highly correlated at \( r = .87 \).

**Results and discussion**

The 4-month-olds differentiated between incongruent word–object pairings (i.e., ‘bubu’ with an angular shape or ‘kiki’ with a curvy shape) and congruent word–object pairings (i.e., ‘bubu’ with a curvy shape or ‘kiki’ with an angular shape). Infants looked longer at incongruent pairings (\( M = 16.8 \text{ s}, SE = 2.3 \)) than at congruent pairings (\( M = 13.0 \text{ s}, SE = 2.0 \)), \( t(11) = 2.77, p < .02, r = .64 \), paired-samples \( t \) test (see Fig. 2, and see Table 1 for mean infant looking times by stimulus characteristics in Experiment 1). Of the 12 infants, 10 looked longer at incongruent trials, \( p < .05 \), using a binomial test. There were no effects of individual sounds or shapes as measured by a two-factor analysis of variance (ANOVA) with shape (curvy or angular) and sound (‘bubu’ or ‘kiki’) (all \( p \) s = ns). There was no interaction between congruency and block in a repeated-measures ANOVA, \( F(3,21) = 0.47, p = .71 \), indicating that the looking time difference between congruent and incongruent trials was not due to learning during the experiment. See Table 2 for a summary of the looking times by block and congruency. There was a significant effect of block, \( F(3,21) = 5.74, p < .01 \), partial \( \eta^2 = .45 \), because infants’ looking times decreased as the experiment progressed.

These results indicate that infants may have some of the same sound–shape associations as adults and toddlers even before they have learned reliable word–object mappings. This pattern of results differs from that of a prior study showing that 4-month-olds look longer at congruent trials (matching [o] or [a] to large objects and [i] or [e] to small objects) than at incongruent trials (Peña et al., 2011). However, this difference may stem from significant methodological differences between the two studies in the number of stimuli presented and in the mode of presentation. The previous study presented six different sounds with four different shapes presented in four different colors, whereas the current study presented two sounds with two shapes. Infants presented with more variables and more complex stimuli tend to look longer at relatively familiar or congruent pairings, whereas infants presented with simpler stimuli tend to look longer at relatively novel or incongruent pairings (Hunter & Ames, 1988). In addition, in the previous study, the two objects were presented simultaneously with one sound, which encourages matching behavior (e.g., Patterson & Werker, 2003). In contrast, in the current study, sound–object pairs were presented sequentially, which can lead infants to look longer at inconsistent audiovisual pairings (e.g., Stager & Werker, 1997). However, both studies provide converging evidence that 4-month-olds differentiate between congruent and incongruent sound–object mappings.

**Table 1**

Mean infant looking times (in seconds) by stimulus characteristics in Experiments 1, 2, and 3.

<table>
<thead>
<tr>
<th>Combination of vowels and consonants</th>
<th>Curvy</th>
<th>Angular</th>
</tr>
</thead>
<tbody>
<tr>
<td>Congruent</td>
<td>13.3 (1.9)</td>
<td>17.6 (2.9)</td>
</tr>
<tr>
<td>Incongruent</td>
<td>12.8 (2.4)</td>
<td>13.6 (1.9)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Vowels</th>
<th>Curvy</th>
<th>Angular</th>
</tr>
</thead>
<tbody>
<tr>
<td>Congruent</td>
<td>12.8 (1.8)</td>
<td>12.1 (2.1)</td>
</tr>
<tr>
<td>Incongruent</td>
<td>11.1 (2.1)</td>
<td>11.9 (1.9)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Consonants</th>
<th>Curvy</th>
<th>Angular</th>
</tr>
</thead>
<tbody>
<tr>
<td>Congruent</td>
<td>9.7 (1.5)</td>
<td>6.9 (1.1)</td>
</tr>
<tr>
<td>Incongruent</td>
<td>7.2 (0.9)</td>
<td>8.5 (1.1)</td>
</tr>
</tbody>
</table>

**Note.** Standard errors are in parentheses.

Please cite this article in press as: Ozturk, O., et al. Sound symbolism in infancy: Evidence for sound–shape cross-modal correspondences in 4-month-olds. *Journal of Experimental Child Psychology* (2012). [http://dx.doi.org/10.1016/j.jecp.2012.05.004](http://dx.doi.org/10.1016/j.jecp.2012.05.004)
Experiment 2

Infants in Experiment 1 showed the same sound–shape mappings as adults and toddlers. However, the words tested in Experiment 1 varied on both vowels (u/i) and consonants (b/k); therefore, it is unclear which segments of the words infants use to make the mappings—consonants, vowels, or a combination of vowels and consonants.

In Experiment 2, we examined whether infant mappings are based on vowels. Vowels are generally longer in duration and have more intensity than consonants (Lagefoged & Johnson, 2010). Moreover, infants can map different vowels onto small and large objects reliably (Peña et al., 2011). Thus, infants may be more likely to use information from the vowels than from the consonants to map sounds to shapes. In Experiment 2, we presented infants with the same two shapes but now did so with two words that differed only in their vowels: ‘kiki’ and ‘kuku’.

Method

Participants

The participants were 12 full-term 4-month-olds (mean age = 4 months 8 days, range = 3 months 22 days to 4 months 23 days, 7 girls and 5 boys). Infants were recruited in the same way as in Experiment 1. Of the infants’ parents, 1 had completed some high school, 2 had completed some college, 11 had completed college, 8 had completed graduate degrees, and 2 declined to answer. Of the 12 infants, 8 were White, 2 were of mixed ethnicity, 1 was Hispanic, and 1 infant’s parent declined to answer. An additional 10 infants were tested but excluded from the analysis because of fussiness (6), inattentiveness (1), or experimenter error (3).

Stimuli

Auditory stimuli. The stimuli for Experiment 2 were recorded during the same session as the stimuli for Experiment 1. The tokens for each word were matched on duration (‘kuku’: M = 435 ms, range = 433–439; ‘kiki’: M = 434 ms, range = 431–437), amplitude (all tokens were 70 dB), and pitch (‘kuku’: M = 232 Hz, range = 231–233; ‘kiki’: M = 232 Hz, range = 231–232). We selected /k/ as the carrier vowel rather than /b/ to limit the variation in rounding information in the vowel and to reduce any coarticulation of rounding from the bilabial consonant /b/ on the vowel. We created 40-s strings in the same way as in Experiment 1.

Visual stimuli. We used the same stimuli as in Experiment 1.

Procedure

The procedure was identical to that in Experiment 1.

Results and discussion

Infants looked equally at incongruent pairings (M = 12.3 s, SE = 1.8) and congruent pairings (M = 11.6 s, SE = 1.8), t(11) = 0.64, p = .54, r = .19 (see Fig. 2, and see Table 1 for mean infant looking times by stimulus characteristics in Experiment 2). Of the 12 infants, 6 looked longer at the congruent trials, p = 1.00, using a binomial test. There was no effect of individual sounds or shape as measured by

| Mean infant looking times (in seconds) in Experiment 1 for each of the four blocks. |
|---------------------------------|-----------------|
| Congruent                       | Incongruent     |
| Block 1 18.6 (3.5)              | 21.4 (3.5)      |
| Block 2 13.6 (2.5)              | 18.0 (3.2)      |
| Block 3 10.9 (2.6)              | 13.9 (2.9)      |
| Block 4 5.7 (1.5)               | 8.9 (2.5)       |

Note. Standard errors are in parentheses.
a two-factor ANOVA with shape (curvy or angular) and sound (‘kuku’ or ‘kiki’) (all ps = ns). Even though infants in prior studies systematically matched vowels to the size of objects (Peña et al., 2011), we found no evidence that infants matched vowels alone to the shape of objects.

**Experiment 3**

Vowel information was not sufficient for infants to match words and shapes. In Experiment 3, we examined whether infants use the consonant information to map words and shapes. Adults prefer using consonants to vowels when making sound–shape pairings (Nielsen & Rendall, 2011). Infants’ use of consonants to make label–shape mappings would be consistent with the role that consonants have been proposed to play in lexical processing, as compared with the role of vowels in indexical, prosodic, and grammatical processing (Bonatti, Peña, Nespor, & Mehler, 2005; Nespor, Peña, & Mehler, 2003). We presented infants with the same two shapes but now did so with two words that differed only in their consonants: ‘bubu’ and ‘kuku’.

**Method**

**Participants**

The participants were 12 full-term 4-month-olds (mean age = 4 months 9 days, range = 3 months 27 days to 4 months 19 days, 7 girls and 5 boys). Of the infants’ parents, 1 had completed some high school, 4 had completed some college, 9 had completed college, 7 had completed graduate degrees, and 3 declined to answer. Of the 12 infants, 5 were Hispanic, 4 were White, 2 were of mixed ethnicity, and 1 infant’s parents declined to answer. An additional 4 infants were tested but excluded from the analysis because of fussiness (1), parental interference (1), or experimenter error (2).

**Stimuli**

*Auditory stimuli.* The stimuli for Experiment 3 were recorded at the same session as the stimuli for Experiments 1 and 2. The tokens for each word were matched on duration (‘kuku’: $M = 398$ ms, range = 387–409; ‘bubu’: $M = 398$ ms, range = 390–407), amplitude (all tokens were 70 dB), and pitch (‘kuku’: $M = 231$ Hz, range = 228–238; ‘bubu’: $M = 230$ Hz, range = 228–232). We selected /u/ rather than /i/ as the carrier vowel because velars such as /k/ have especially low resistance to coarticulation from /i/ (Fowler & Brancazio, 2000). We created 40 s strings in the same way as in Experiment 1.
Visual stimuli. We used the same stimuli as in Experiment 1.

Procedure
The procedure was identical to that in Experiment 1.

Results and discussion
Infants looked equally at incongruent pairings (M = 7.5 s, SE = 1.0) and congruent pairings (M = 8.5 s, SE = 0.9), t(11) = 0.87, p = .40, r = .25 (see Fig. 2, and see Table 1 for mean infant looking times by stimulus characteristics in Experiment 3). Of the 12 infants, 7 looked longer at congruent trials (p = .77) using a binomial test. There were no effects of individual sounds or shape as measured by a two-factor ANOVA with shape (curvy or angular) and sound (‘bubu’ or ‘kuku’) (all ps = ns). Despite the importance of consonants in lexical processing (Bonatti et al., 2005; Nespor et al., 2003), we found no evidence that infants matched consonants alone to the shape of objects.

Comparison of infants’ performance across the three infant experiments
To compare the three infant experiments, we derived each infant’s relative looking time to incongruent trials by calculating a difference score between incongruent and congruent looking times (Incongruent Looking Time – Congruent Looking Time). A one-way ANOVA comparing difference scores across the three experiments revealed a reliable effect of experiment, F(2,33) = 4.66, p = .02, η² = .22. Post hoc least-squares differences tests confirmed that infants looked relatively longer to incongruent trials in Experiment 1, which combined both vowels and consonants, than in Experiment 2, which compared vowels (p = .02, r = .45), and in Experiment 3, which compared consonants (p = .01, r = .48). Infants differentiated between congruent and incongruent trials only when the labels differed on both consonants and vowels (see Fig. 2).

Experiment 4
Previous work examining the features of speech sounds used by adults to make sound–object mappings found that participants were more likely to respond based on consonant quality than on vowel quality when the two made opposite predictions (Nielsen & Rendall, 2011). The current experiments with infants used different words (‘bubu’, ‘kiki’, and ‘kuku’) from those used in previous studies. In Experiment 4, we tested adults on these new materials to confirm that adults generalize sound–shape mappings to these words. Furthermore, we examined adults’ use of vowels or consonants in making systematic mappings.

Method
Participants
The participants were 10 undergraduate students (9 women and 1 man) either enrolled at New York University or participating in a summer internship program at the university. All of the participants were native English speakers and between 18 and 22 years old. An additional 3 participants’ data were not included in the analysis because they were not native English speakers (2) or did not follow task instructions (1).

Stimuli
Auditory stimuli. These were the same as those used in the previous experiments with the addition of ‘bibì’. The tokens were adjusted using Praat (Boersma & Weenink, 2010) to be similar in duration (M = 425 ms, range = 423–438), pitch (M = 227 Hz, range = 225–230), and intensity (all 70 dB).

Visual stimuli. These were the same shapes as those used in the previous experiments except that we included more colors to encourage independent judgments for each trial: red, light pink, bright pink, maroon, brown, rust, orange, mustard, marigold, yellow, bright green, seafoam green, emerald green.
olive green, teal, turquoise, French blue, royal blue, light gray, dark gray, gray–green, blue–purple, and purple.

**Procedure**

**One-label task.** Participants sat at a desktop computer wearing Extreme Isolation Ex-25 noise-canceling headphones. They were presented with two shapes: a curvy shape and an angular shape of the same color, one on each side of the computer screen. The shapes remained consistent across all of the trials but varied in color between trials. The two objects presented in any given trial were of the same color. Below these shapes was a teardrop (half curvy and half angular). Participants were instructed to click on the teardrop to hear a word and then to record whether the word was a better match for the shape on the left or the right. Two trials for each of the nonwords used in the infant experiments ('bubu', 'kiki', or 'kuku'), with the addition of 'bibi', were presented for a total of eight experimental trials.3

**Two-label task.** Participants sat at a desktop computer wearing Extreme Isolation Ex-25 noise-canceling headphones. They were presented with the same two shapes as in the one-label task but with an additional teardrop above the objects. The second teardrop played a different word when clicked. Participants were instructed to click on the two teardrops to hear two words and then to pair each of the words with one of the two objects. This procedure allowed the isolation of specific features for word–object pairings. For example, a trial with ‘kiki’ and ‘kuku’ measured whether participants could use vowel information alone to make consistent pairings without conflicting or supporting consonant information. There was a total of six test trials—one trial for each permutation of /b/ and /k/ combined with /i/ and /u/ (kiki/kuku, bibi/bubu, kuku/bubu/, kiki/bibi, bibi/kuku, and bubu/kiki).

**Results and discussion**

**One-label task**

Trials were scored according to whether participants used a vowel (e.g., pairing /i/ with the angular shape), a consonant (e.g., pairing /k/ with the angular shape), or a combination of the two (e.g., pairing /kiki/ with the angular shape). Participants had the highest score using the combinations (M = 95%), followed by consonants (M = 88%). Scores on trials comparing vowels (M = 63%) were significantly lower than either consonant trials, t(9) = 4.58, p < .01, r = .84, or combination trials, t(9) = 9.80, p < .01, r = .96.

**Two-label task**

Trials were scored according to whether participants assigned the words to the congruent shapes (/i/ or /k/ with the spiky shape and /u/ or /b/ with the curvy shape). Accuracy was equivalent between trials when only vowel information distinguished the words (kiki/kuku or bibi/bubu, M = 90%), when only consonant information distinguished the words (kiki/bibi or kuku/bubu, M = 90%), and when the consonant–vowel combination distinguished the words (bubu/kiki, M = 100%). However, when words contained conflicting vowel and consonant information (kuku/bibi), 8 of the 10 participants matched the words according to consonant information (p < .05) using a binomial test.

Adults’ use of consonants and vowels differed from results of some previous studies in which adults used only consonant information to match shapes and sounds (Nielsen & Rendall, 2011). This may be because previous studies used different materials with multiple consonants and vowels—‘takete’ and ‘maluma’ rather than ‘bubu’/‘buba’ and ‘kiki’. However, the difference in findings is more likely due to the different tasks. The task used by Nielsen and Rendall (2011) is more similar to the one-label task, in which participants need to choose whether to use vowels or consonants to match the shapes. The one-label task results are consistent with Nielsen and Rendall’s findings in showing that participants will use consonants instead of vowels. However, Nielsen and Rendall did not contrast labels differing

---

3 Participants were also presented with trials where they heard the non-English vowels (/i/ and /y/) combined with both /b/ and /k/, but these trials were not analyzed.
only on vowels; thus, their results cannot speak to whether participants can use vowel information at all. Our results suggest that participants prefer to match using consonant information but can use vowels when no consonant information is available (see also Monaghan et al., 2012).

General discussion

We predicted that if 4-month-old infants have the same sound–shape mapping biases as adults and toddlers, they would differentiate between congruent trials (pairing the angular shape with ‘kiki’ or the curvy shape with ‘bubu’) and incongruent trials (pairing the curvy shape with ‘kiki’ or the angular shape with ‘bubu’). Infants looked longer at incongruent pairings than at congruent pairings. This mapping appears to be based on a combination of vowels and consonants given that neither vowels alone nor consonants alone elicited systematic mapping. Our results indicate that adult sound–shape mapping biases have their origins during infancy and might not result uniquely from language exposure creating statistical regularities in the lexicon.

Adults in our study, like infants, used a combination of consonant and vowel information to match the labels they heard with the shapes they saw. Their better performance with a combination of consonants and vowels rather than either vowels or consonants alone may be because the combination of vowels and consonants contained more sound symbolic phonemes, thereby inducing a stronger sound symbolism effect (Thompson & Estes, 2011). However, this was not the only strategy that was available to them. Adults, unlike infants, were also able to use consonant information alone and vowel information alone to match the labels to the shapes, albeit less frequently than the consonant–vowel combination. When vowels and consonants were put in conflict, adults used consonants more often than vowels. Adults’ primary use of consonants is consistent with the role that consonants have been proposed to play in lexical processing, as compared with the role of vowels in indexical, prosodic, and grammatical processing (Bonatti et al., 2005; Nespor et al., 2003). Specifically, consonants and vowels are processed differently: adults extract statistical regularities from consonants but not from vowels (Bonatti et al., 2005; Toro, Shukla, Nespor, & Endress, 2008; but see Newport & Aslin, 2004), whereas rule extraction is privileged over vowels but not over consonants for adults and infants (Bonatti et al., 2005; Hochmann, Benavides-Varela, Nespor, & Mehler, 2011; Pons & Toro, 2010; Toro et al., 2008).

Where do these biases for systematically mapping sounds to shapes come from? One possibility is that sound symbolism is related to cross-modal synesthetic tendencies (e.g., Ramachandran & Hubbard, 2001; Spector & Maurer, 2009). An adult with synesthesia experiences additional involuntary sensations in one sensory modality when another sensory modality is stimulated (Cytowic & Eagleman, 2009; Cytowic & Wood, 1982). Synesthetic associations are stronger in younger infants than in older infants (Wagner & Dobkins, 2011), indicating that these associations may play a larger role in the perceptions of prelinguistic infants than of toddlers or adults. However, because synesthesia occurs in only 4 to 5% of the adult population (Sinner & Ward, 2006), whereas 90% to 98% of adults show systematic sound–shape mapping biases (Köhler, 1947), our results are unlikely to be explained by adult synesthesia. At the same time, the human proclivity for systematically mapping sounds to shapes is also consistent with data from nonhuman primates showing biased cross-modal correspondences in chimpanzees (Ludwig, Adachi, & Matsuzawa, 2011). Thus, sound symbolism may be an intrinsic feature of primate sensory systems.

Another potential explanation for the origins of cross-modal biases comes from theories claiming links between perception and production mechanisms (Calvert & Campbell, 2003; Keysers et al., 2003; Kohler et al., 2002; Liberman & Mattingly, 1985; Wilson, Saygin, Sereno, & Iacoboni, 2004) and is supported by studies with individuals with autism spectrum disorder and patients with brain lesions. These individuals do not show as systematic a bias for sound–shape mappings as typical individuals do. Where typically developing individuals agree with the standard result 90% to 98% of the time, individuals with autism or lesions to the angular gyrus agree only 56% of the time (Ramachandran, Azoulay, Stone, Srinivasan, & Bijoy, 2005). Children with autism spectrum disorder also do not show the sound–shape mapping biases that neurologically typical adults and children show (Oberman & Ramachandran, 2008). Individuals with autism spectrum disorder or brain lesions are proposed to
have a dysfunction in the mirror neuron system linking perception and production that prevents the systematic bias for sound–shape mappings that typical adults show (Altschuler et al., 2000; Nishitani, Avikainen, & Hari, 2004; Oberman et al., 2005; Theoret et al., 2005).

The current results provide evidence that prelinguistic infants show adult-like nonarbitrary mapping biases between speech labels and objects. Infants in our experiments showed these biases only when the two labels differed in both consonants and vowels. Our results differ from those of prior studies in which vowels alone were sufficient to convey object size to 4-month-olds (Peña et al., 2011). The use of vowels alone in prior studies may be grounded in observable physical relationships between the pitch of vowels and the size of objects built through participants’ prior multisensory experiences. The relationship between vowel quality and object shape, in contrast, might not be as easily observable, consistent with findings that consonants, and not vowels, are preferred in adult sound–shape mappings (Experiment 4 in the current study; see also Nielsen & Rendall, 2011). It must be noted that our experiments used only two consonants (/b/ and /k/) and two vowels (/i/ and /u/). Although these data are consistent with previous infant studies linking vowel and object size (Peña et al., 2011) and pitch and object shape (Walker et al., 2010), further research with different consonants and vowels is needed to understand what may drive infants’ sound–shape mapping biases.

Our perceptual system must combine or segregate concurrent sensory inputs from different modalities to quickly and accurately detect objects or events and to choose appropriate responses (Evans & Treisman, 2010). In a world full of multisensory experiences, developing infants must be able to integrate multisensory inputs (e.g., Bahrick, Lickliter, & Flom, 2004; Lewkowicz, 2000). The early biases for cross-modal correspondences during infancy provide a basis for processing the multisensory nature of our daily experiences.

Acknowledgments

This research was supported by New York University (NYU) research funds to Athena Vouloumanos. We thank Ilana Shanks and the members of the NYU Infant Cognition and Communication Lab for their help in running and coding the studies and thank the parents and infants who participated in this research.

References


Please cite this article in press as: Ozturk, O., & al. Sound symbolism in infancy: Evidence for sound–shape cross-modal correspondences in 4-month-olds. Journal of Experimental Child Psychology (2012), http://dx.doi.org/10.1016/j.jecp.2012.05.004


Please cite this article in press as: Ozturk, O., et al. Sound symbolism in infancy: Evidence for sound–shape cross-modal correspondences in 4-month-olds. Journal of Experimental Child Psychology (2012), http://dx.doi.org/10.1016/j.jecp.2012.05.004