Do 14-17-Month-Old Infants Use Iconic Cues to Interpret Words?

Suzanne Aussems (s.aussems.1@warwick.ac.uk)
Department of Psychology, University of Warwick
Coventry, CV4 7AL, United Kingdom

Lottie Devey Smith (cds224@exeter.ac.uk)
School of Education, University of Exeter
Exeter, EX1 2LU, United Kingdom

Sotaro Kita (s.kita@warwick.ac.uk)
Department of Psychology, University of Warwick
Coventry, CV4 7AL, United Kingdom

Abstract

This study investigated whether infants use iconicity in speech and gesture to interpret words. Thirty-six 14-17-month-old infants participated in a preferential looking task in which they heard a spoken non-word (e.g., *"zudzud"*) while observing a small and a large object (e.g., a small and a large square). All infants were presented with an iconic cue for object size (small or large) in 1) the pitch of the spoken non-word (high vs. low), 2) in gesture (small or large), or 3) congruently in both pitch and gesture (e.g., a high pitch and a small gesture indicating a small square). Infants did not show a preference for congruently sized objects in any iconic cue condition. Bayes Factor analyses supported the null hypotheses. In conclusion, we found no evidence that infants link the pitch of spoken non-words, or the iconic gestures accompanying those spoken non-words, to object size.

Keywords: Iconicity; Sound symbolism; Pitch, Size; Gesture

Introduction

A traditional linguistic view is that language is arbitrary, meaning there is no natural motivation between the sound of a word and its meaning (de Saussure, 1983). However, research on sound symbolism provides evidence for non-arbitrary form-meaning mappings in vocabulary (Monaghan et al., 2014; Ohala, 1984; Perry et al., 2018). In English, words like "shimmer", "buzz" and "hiss" comprise sounds that iconically depict their meanings (Ohala, 1984). Cross-linguistic studies also report sound-meaning systematicity in many languages (Brown, 1958; Monaghan et al., 2014). Recently, iconicity has been considered an important design feature that helps infants break into the language system (Imai & Kita, 2014; Perniss & Vigliocco, 2014). Corpus analyses confirm that children's early vocabulary is more iconic, reducing as development progresses (Monaghan et al., 2014; Perry et al., 2018). Iconicity likely plays a bigger role in early language acquisition than was previously thought. Therefore, the current study investigated whether infants use iconic cues to interpret word meanings.

Segmental and Suprasegmental Sound Symbolism

There are two types of sound symbolism—segmental and suprasegmental. Segmental sound symbolism refers to individual speech sounds mapping to meaning. For example, Köhler (1947) showed that people map words with rounded sounds (e.g. *"maluma"*) to curvy shapes and words with unrounded sounds (e.g. *"takete"*) to spiky shapes. This effect is robust across languages and age groups (Fort et al., 2018; Maurer et al., 2006).

Suprasegmental sound symbolism refers to speech properties like pitch, amplitude, and duration conveying meaning. So, these are sound-symbolic effects that transcend individual words. For example, words spoken faster can indicate fast motion, and words spoken slower can indicate slow motion (Shintel et al., 2006). Another example is pitch conveying size, which is the focus of this study. Across languages, high pitch indicates smallness, while low pitch indicates largeness (Ohala, 1984). Investigating the cross-modal correspondence between pitch and size helps us understand how infants combine auditory and visual inputs.

Cross-Modal Pitch-Size Correspondences

Two lines of empirical work demonstrate this pitch-size cross-modal mapping. First, studies show adults map pure tone pitches to object sizes. For example, adults made faster size judgments about disks when tones were congruent vs incongruent with size (Gallace & Spence, 2006). And they categorized combinations of pitches and circles faster when these pairings were size-congruent (high pitch + small circle; low pitch + large circle) (Parise & Spence, 2012).

Second, studies using speech pitches showed adults and children map low pitch adjectives to large objects and high pitch to small objects (Herold et al., 2011; Nygaard et al., 2009; Reinisch et al., 2013). Furthermore, adults rated a sandwich as larger in size when a spoken advertisement for the sandwich used lower pitch (Lowe & Haws, 2017).

However, evidence for this pitch-size correspondence in infants is limited. While studies showed that 6-month-olds (Fernández-Prieto et al., 2015) and 30-35-month-olds (Mondloch & Maurer, 2004) map pure tone pitches to sizes, to our knowledge, no study has shown this with speech stimuli. Therefore, the first research question of this study is whether 14-17-month-old infants can use pitch in speech sounds to interpret novel word meanings? If so, then this would provide evidence for a supporting role of suprasegmental sound symbolism in language development.

In L. K. Samuelson, S. L. Frank, M. Toneva, A. Mackey, & E. Hazeltine (Eds.), *Proceedings of the 46th Annual Conference of the Cognitive Science Society.*

When does Iconic Gesture Comprehension Start?

Iconicity is not only present in speech but also in gesture (Aussems & Kita, 2019). Iconic gestures depict meaning visually through hand movements (McNeill, 1992). For example, people may move their hands far apart to indicate the size of a fish they caught. It is unclear when children start to comprehend iconicity in gestures. Some research suggests that children do not understand iconic gestures before age 3 (Stanfield et al., 2014). However, other studies demonstrate comprehension by age 2 (Namy, 2008; Tomasello et al., 1999) and verb learning and generalization via iconic gesture by age 3 (Aussems & Kita, 2021). To date, no study has shown iconic gesture comprehension in infants under 18 months. Previous research is limited because it used only explicit measures (e.g., asking children to point at the referent of a gesture). By using an implicit measure, this study investigates if 14-17-month-old infants already show iconic gesture comprehension. If so, then this would show earlier iconic gesture comprehension than previously reported.

Bootstrapping Iconic Gesture Comprehension

Given the available empirical evidence available, it appears that infants' sensitivity to sound symbolism appears before their sensitivity to iconic gestures. However, it is unclear if their emerging sensitivities to iconicity are developmentally interrelated, or unrelated cognitive processes. If interrelated, iconic speech comprehension may bootstrap iconic gesture comprehension. Therefore, the third research question of this study is whether iconic speech helps infants to interpret iconic gestures. If so, then this would suggest sound symbolism plays an additional developmental role, providing a multimodal footing for iconic mappings.

The Current Study

This study used a preferential looking task, with 14-17-month-old infants viewing small/large object pairs (e.g., a large and a small square), while hearing spoken non-words across three iconic cue conditions: 1) *speech*, 2) *gesture*, and 3) *multimodal*. In the *speech* condition, the spoken nonword cued the large object via low pitch and the small object via high pitch. In the *gesture* condition, the co-speech gesture cued the large object via hands moving far apart and the small object via hands moving close together. In the *multimodal* condition, a combination of speech and gesture cues congruently indicated the same object size (small or large).

We predict that infants' preferences for word referents of a certain size will differ by iconic cue (H1). Specifically, we predict that infants will prefer congruent sized objects in all conditions (H2-H4), but that this preference will be strongest in the multimodal condition (H5). This is because speech cues may bootstrap infants' understanding of gesture cues, leading to an extra-additive effect.

Positive evidence would provide new data on the potential supporting role of suprasegmental sound symbolism in early word learning, early capacity for iconic gesture comprehension, and a developmental link between emerging sensitivities to iconicity across communicative domains.

Method

Pre-Registration & Open Materials & Data

The hypotheses, methods, materials, and analyses of this study were pre-registered via the Open Science Framework (OSF) prior to data collection and can be accessed via https://osf.io/k89mp. The anonymized numerical (raw) data and analysis scripts are documented on OSF and can be accessed via: https://osf.io/b65sd/.

Design

The experiment had a within-subject design in which iconic cue was manipulated. This independent variable had three levels: *speech*, *gesture*, and *multimodal*. All infants completed the same 8 test trials in each iconic cue condition, resulting in a total of 24 test trials. The dependent variable was the average proportion of looking time towards the target objects, which were the objects of a size congruent with one of the iconic cues or multimodal cues.

Power analysis

We ran two power analyses prior to data collection to determine the sample size for the study. Specifically, we ran one power analysis for the repeated-measures ANOVA and one for the one-sample t-tests in G*Power version 3.1.9.2 (Faul et al., 2007) to calculate the sample size needed to detect an effect with a power of 0.90 and an error probability (a) of 0.05. We chose the default settings in G*Power for a medium-sized effect (f = 0.25 for the ANOVA and d = 0.50 for the t-tests, respectively). These power analyses resulted in a minimum sample size of 36 infants.

Participants

Infants were recruited via a database of families who expressed interest in taking part in developmental research. The final sample included 36 typically developing infants (17 girls) between 14 and 17 months old (M = 16.21, SD = 1.15). There were seven 14-month-olds (2 girls), eight 15-month-olds (4 girls), eleven 16-month-olds (6 girls) and ten 17-month-olds (5 girls). Infants were exposed to English for more than 75% of the time at home. Infants received a small toy for taking part. An additional 12 infants were tested but excluded from the analysis because they did not look at the objects in more than half of the trials and in at least three trials per condition (N = 10 infants), or they were too old on the day of their testing appointment (N = 2 infants).

Apparatus

Testing took place in a sound-attenuated testing booth with an adjacent control room. The experiment was run on a Mac computer using Habit (version 1.0) software. The visual and auditory stimulus materials were combined in the Habit software to create the experiment. The visual stimuli were displayed on a 40-inch Phillips TV monitor and the auditory stimuli were played through two Pulse audio speakers. A digital Canon Legria HF R56 video camera was located

below the TV monitor, at the infant's eye height. The distance between the infant and the camera lens was approximately 100 cm. To minimize distractions, the testing area was covered in black curtains, which surrounded the monitor and hid the speakers and camera from view. The camera sent live footage of the session from the testing area to a control area so that the researcher could monitor the session. Testing sessions were recorded for later offline coding.

Materials

There were five pre-test trials to familiarize infants with the actress, pitches, and gestures from the main experiment task. One pre-test trial showed a video of an actress waving and saying "Hi! Hello!"; two pre-test trials showed videos in which the actress produced small, medium, and large iconic gestures (without any audio); and two pre-test trials played audio clips of spoken non-words (which did not appear in the test trials) at low, neutral, and high pitches (without a visual).

The 24 test trials combined video and audio recordings. The 14-second videos featured static large-small object pairs (8 objects in total) as well as an actress who produced a large, neutral, or small gesture after 3, 7, and 10 seconds. Eight nonsense words were recorded at low, medium, and high pitches and combined with the videos at 3, 7, and 10 seconds. The actress wore a surgical mask, so it was possible to combine each audio clip with the same gesture video without a mismatch between her lip movements and what she said.

Procedure

Infants were placed in a highchair in front of the TV monitor, with their caregiver sitting in a chair slightly behind them. Caregivers were instructed not to interfere during the task and listened to music via noise-cancelling headphones.

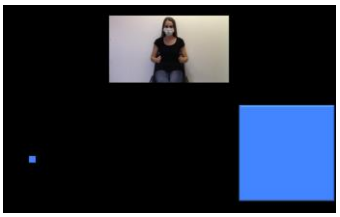
In the pre-test phase, infants first saw the 5 pre-test trials (actress waving, gesture videos, audio recordings) in a counterbalanced order. In the test phase, infants saw each of the 24 test trials combining video and audio recordings. The trials were 14 seconds long and the task took approximately 5 minutes. In between trials, an attention getter (a color-changing circle) appeared in the center of the screen to draw infants' attention. The researcher, who viewed the live feed of the session from the control area, only started a trial if the infant looked at the attention getter. Infants' gaze was coded in real time and trials were repeated if the infant looked away.

Counterbalancing & Randomization

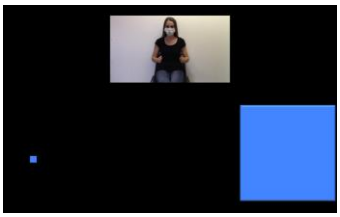
Each participant completed 24 test trials in three blocks of eight trials, with each block containing the same number of trials in each condition and each object appearing only once in each block. Thus, each object appeared three times in the task, but in a different iconic cue condition. The size of the target object (small or large) was counterbalanced, as well as the order of the blocks (1, 2, and 3). This resulted in six versions of the experiment. For a given participant, each object appeared equally often in each of the three conditions, the target and distractor objects appeared equally often on the left and right sides of the screen, and the small and large objects appeared equally often on the left and right sides. The same spoken non-word was always used for the same object (*"datdat"* = circle, *"dutdut"* = cross, *"satsat"* = diamond, *"sutsut"* = hexagon, *"tadtad"* = spikey circle, *"tudtud"* = square, *"zadzad"* = triangle, *"zudzud"* = wavy rectangle). Multisyllabic words make pitch information clearer than monosyllabic words (i.e., they contain more vowels that can carry pitch information). And reduplication within words simplifies the segmental information that infants need to process, so that they can focus better on pitch. Previous research has also shown that infants learn reduplicated words more easily (Ota & Skarabela, 2016).

Figure 1: Examples of test trials in three iconic cue conditions

Speech condition

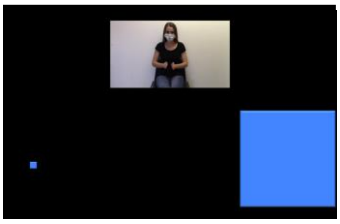


High pitch "*tudtud*" with neutral gesture

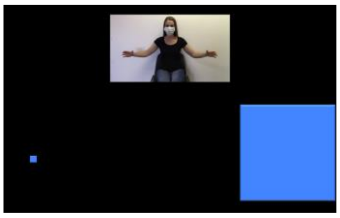


Low pitch "*tudtud*" with neutral gesture

Gesture condition

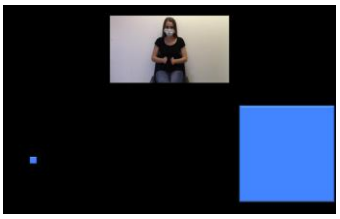


Small gesture with neutral pitch *"tudtud"*

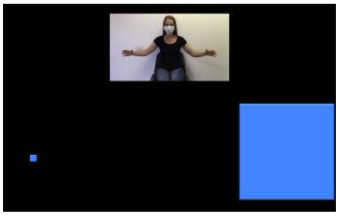


Large gesture with neutral pitch *"tudtud"*

Multimodal condition



High pitch *"tudtud"* with small gesture



Low pitch *"tudtud"* with large gesture

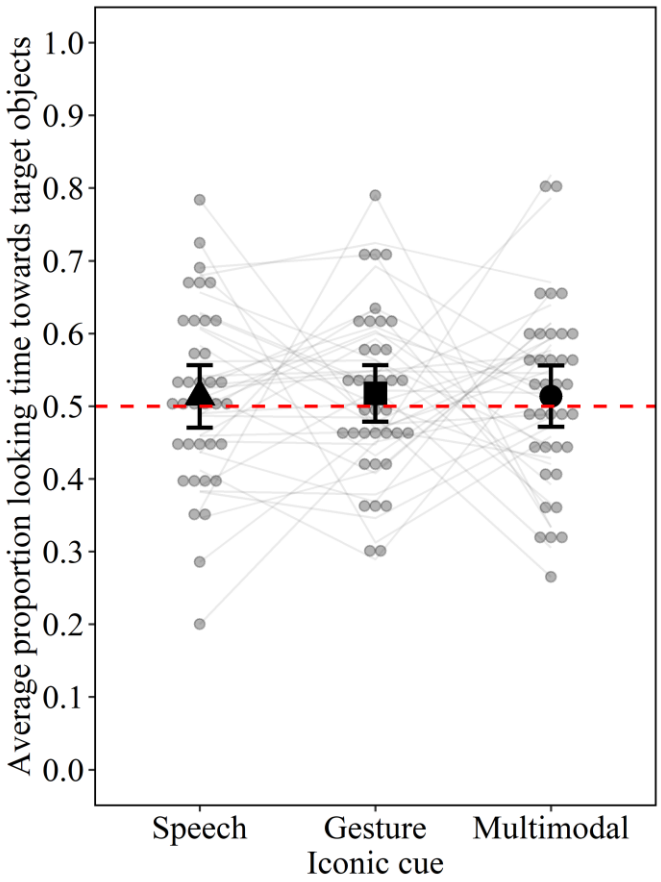
Coding Looks & Inter-Rater Reliability

The first human coder used Eudico Linguistic Annotator software (ELAN, version 5.3) (Lausberg & Sloetjes, 2009) to annotate the infants' looks frame-by-frame (one frame = 40 ms). The coder used look directions for the left target, right target, left distractor, right distractor, center, caregiver, and elsewhere. The window of analysis started 367 milliseconds (ms) from the onset of the speech and gesture stimuli to allow the iconic cues to unfold, and it ended when the trial ended. The second human coder who was naive to the study's purpose and hypotheses independently annotated the looks of 12 infants (33%). Inter-rater reliability was calculated based on 1859 overlapping annotations using Spearman's rank-

order correlation coefficient for look duration in milliseconds and Cohen's Kappa for look direction. Look duration (ms) was positively and significantly correlated between the two independent coders, r (1859) = .846, z = 37.45, p < .001, 95% CI [.801, .890], indicating excellent reliability. There was also excellent agreement (99.1%) and inter-rater reliability, κ = .989, p < .001, 95% CI [.984, .994], between the categories of look directions of the two coders.

Results

Figure 2: Infants' preference for target objects by iconic cue



Note. Average proportion looking time towards target objects (*y*-axis) by iconic cue (*x*-axis). Shapes represent the means in each condition. Error bars represent 95% confidence intervals around the means. Grey dots represent individual performances and grey lines link performances of the same infants across conditions. Red horizontal dashed line represents the chance level (.50).

Trial Exclusions

There were 864 trials for 36 infants (36 x 24 = 864). None of the participants showed an extreme size bias (i.e., looked towards either the small or large objects more than 80% of the time in more than 90% of the trials) or an extreme side bias (i.e., looked towards objects on either the left or the right side of the screen for more than 80% of the time in more than 90% of the trials). Due to a human error, 18 participants received an extra trial in the *gesture* condition instead of the *multimodal* condition. These extra trials were included in the analysis. A total of 131 trials were excluded due to the infant not looking at the objects in the window of analysis or the sum of the looking time towards the two objects being less than 5% in the window of analysis (N = 106), caregiver interference (e.g., the caregiver pointed at the screen) (N = 7), noncompliance (e.g., the infant did not want to sit in the highchair) (N = 8), and fussing or crying such that both eyes were not visible (e.g., the infant rubbed their eyes) (N = 10). We analyzed the data of the remaining 733 trials (250 in the *speech* condition, 261 in the *gesture* condition, and 222 in the *multimodal* condition).

Does Infants' Size Preference Differ by Iconic Cue?

To test H1, we conducted a one-way repeated-measures ANOVA which revealed no significant difference in average proportions of looking time towards targets between iconic cue conditions, F (2, 70) = 0.01, p = .989, η2 = 0.00. Infants looked towards target objects equally in the *speech* (M = .51, SD = .13), *gesture* (M = .52, SD = .12), and *multimodal* (M = .51, SD = .13) conditions (see Figure 2).

We also calculated Bayes factors to test whether the data are inconclusive, or if the experimental manipulation did not have any effect. A Bayesian one-way repeated-measures ANOVA revealed strong evidence in favor of the null hypothesis; that is, there was no significant difference in the average proportions of looking time towards targets between iconic cue conditions (BF01 = 11.56). This Bayes Factor indicates that the data are 11.6 times more likely under the null hypothesis than under the alternative hypothesis.

Chance Comparisons

Next, to investigate H2, H3, and H4, we compared the average proportions of looking time towards target objects in each iconic cue condition against chance (test value = .50) using one-sample *t*-tests (two-tailed). The average proportion of looking time towards target objects of .51 (SD = .13, 95% CI [.47, .56]) in the *speech* condition was not significantly above chance, t (35) = 0.63, p = .532, Cohen's d = 0.11, 95% CI for Cohen's d [-0.22, 0.43], neither was the average proportion of looking time towards target objects of .52 (SD = .12, 95% CI = [.48, .56]) in the *gesture* condition, t (35) = 0.90, p = .373, Cohen's d = 0.15, 95% CI for Cohen's d [-0.18, 0.48], nor was the average proportion of looking time towards target objects of .51 (SD = .13, 95% CI = [.47, .56]) in the *multimodal* condition, t (35) = 0.66, p = .516, Cohen's d = 0.11, 95% CI for Cohen's d [-0.22, 0.44]. Next, we conducted the same chance comparisons using Bayesian one-sample *t*-tests (two-tailed, test value = .50). These tests showed moderate evidence in favor of the null hypothesis; that is, the average proportion of looking time towards target objects was not significantly different from chance in the *speech* condition (BF01 = 4.64), neither in the *gesture* condition (BF01 = 3.83), nor in the *multimodal* condition (BF01 = 4.57). These Bayes Factors indicate that the data are between 3.8 and 4.6 times more likely under the null hypotheses than under the alternative hypotheses.

Extra-Additive Effect

Finally, to examine H5, we tested whether there was an extra-additive effect of the *multimodal* condition, on the average proportion of looking time towards target objects, compared to the combined effects of the *speech* and *gesture* conditions. A one-sample *t*-test (two-tailed, test value = 0) showed that the average extra-additive score of .02 (*SD* = .24, 95% CI [-.07, .10]) was not significantly different from 0, $t(35) = 0.42$, $p = .675$, Cohen's $d = 0.07$, 95% CI for Cohen's d [-0.26, 0.40]. An equivalent Bayesian one-sample *t*-test (two-tailed, test value = 0) showed moderate evidence in favor of the null hypothesis; that is, the extra-additive effect score was not significantly different from 0 (BF01 = 5.14). This Bayes Factor indicates that our data are 5.1 times more likely under the null hypothesis than under the alternative hypothesis.

General Discussion

This study investigated experimentally whether 14-17-month-old infants use iconic speech and gesture cues to interpret words. Specifically, we examined whether infants showed a preference for objects of a particular size if the object's size was congruent with a single iconic cue or multimodal iconic cues. We evaluated five hypotheses. First, we predicted that infants' preference for objects of a certain size would differ by iconic cue (H1). We also predicted that infants would show a reliable sensitivity to pitch-size cross-modal correspondences in the *speech* (H2), *gesture* (H3), and *multimodal* (H4) conditions. Finally, we predicted that infants would show an extra-additive sensitivity to pitch-size cross-modal correspondences in the *multimodal* condition (H5). The infants in our study showed no preference for small or large objects in any of the iconic cue conditions. Follow-up Bayes Factor analyses indicated moderate to strong evidence for the null hypotheses. This suggests that our data are more consistent with the prediction of no effects than with the prediction of some effects. Although this study does not support infants' early iconic cue comprehension or the bootstrapping role of sound symbolism in iconic gesture comprehension, it remains significant. By pioneering a novel paradigm and testing infants' comprehension of iconicity in different modalities, our study contributes to the literature on the role of iconicity in early language development. Further investigation could reveal the potential role of suprasegmental sound symbolism in early word learning and its link to emerging sensitivities to iconicity across communicative domains.

We believe that our study did not result in a Type II error (false negative). This could happen when the statistical test used is not powerful enough, the effect size is too small, or the sample size is too small. However, we find this scenario unlikely for the following reasons. First, our Bayes Factor analysis provides evidence in favor of the null hypothesis, reinforcing the reliability of our findings. Second, our experiment employed a within-subject design and obtained substantial statistical power: 90% power ($\alpha = 0.05$) to detect a medium-sized effect, with a sample of 36 infants.

We believe that our paradigm effectively captured infants' looking times. The large object was always presented ten times larger than the small object, creating a distinct size difference between the two. We also did not detect an extreme size bias or a side bias in individual infants. Furthermore, the robustness of our data—exhibiting a normal distribution with minimal outliers—supports the notion that our preferential looking paradigm effectively captured infants' looking times. Future studies could explore this paradigm's impact on experimental outcomes and uncover nuances that enhance its utility.

We also acknowledge the need for further investigation to ascertain whether infants can effectively utilize iconic pitch and gesture cues to interpret words. To validate the materials and procedure, future research should include testing both children and adults using this paradigm. If older children and adults do not exhibit the expected effects, it may indicate specific limitations in our paradigm. Conversely, if children and adults demonstrate sensitivity to the iconic pitch and gesture cues employed in our study, we can infer that infants may not yet possess this same sensitivity.

The Pitch Manipulation

The lack of an iconic pitch effect in this study is inconsistent with studies that reported a positive effect of pure tones on infants' (Fernández-Prieto et al., 2015) and children's preferences for objects of a certain size (Mondloch & Maurer, 2004). However, it is important to consider that previous studies used pure tones and our study used spoken-non words as stimuli. The pitch manipulation in our spoken non-words may have been too weak. We varied the pitch of the spoken non-words between trials, which was the only way for infants to interpret this iconic cue—relative to the previous pitch(es) they heard. This is also why we introduced infants to the low, neutral, and high pitches in the pre-test trials before the test trials began. However, it is probably more effective to draw infants' attention to pitch if it rises or falls within a spoken non-word, and thus within a trial, mimicking studies in which whistle tones with an ascending or falling frequency were used as auditory stimuli (e.g., Fernández-Prieto et al., 2015).

Additionally, pitch alone might not exert a significant effect unless it is embedded within a broader acoustic signature. In our study, pitch was the only manipulated acoustic feature and so it was not embedded within a broader acoustic signature. Alternatively, it is plausible that amplitude and duration exert a more pronounced influence on the cross-modal effect observed in prior studies, while pitch plays a secondary role. When mothers produced adjectives like 'big' and 'small' for 2-year-old children, no significant pitch difference was observed, but variations in amplitude and duration were evident (Herold et al., 2011). This preference for other acoustic features – such as amplitude and duration – when emphasizing large and small objects may explain the lack of a pitch effect in the current study. A pivotal question for future research remains: Can infants use suprasegmental cues to interpret word meanings?

The Gesture Manipulation

The lack of an iconic gesture effect in this study is consistent with previous studies, in which infants under 18 months showed no reliable iconic gesture comprehension (Namy, 2008; Stanfield et al., 2014; Tomasello et al., 1999). The infants in our study may not have understood the iconic gestures because these gestures depicted attributes of objects. In a study by Hodges et al. (2018), an experimenter presented 2-, 3-, and 4-year-olds with iconic gestures depicting one of six objects, accompanied by minimally informative speech. Half of the gestures conveyed action information (e.g., "I have this one" + extended index finger moving up and down rapidly as if bouncing a basketball) and the other half conveyed attribute information (e.g., "I have this one" + cupped hands held apart indicating the shape and size of a tennis ball) associated with the objects. After each iconic gesture, the experimenter presented the child with pictures of two objects: a correct match (e.g., basketball) and an incorrect match (e.g., bird) for the object depicted in gesture. The children were then asked to choose one of the pictures ("Which one did I have?"). Children at all ages were able to reliably pick out the match of the iconic gesture that depicted actions, but only 3- and 4-year-olds, and not 2-year-olds, were able to reliably pick out the match of the iconic gestures that depicted attributes of objects. Thus, certain types of iconicity in gesture are understood before others. The gestures in our study depicted object size, and this type of iconicity may have been too difficult to understand for infants. Future studies could investigate if infants understand iconic gestures that depict action information using our preferential looking paradigm.

Furthermore, our iconic gestures had an imaginary component which may have led to null results. Some iconic gestures directly represent the intended meaning. For example, when someone clenches their fist to depict a tennis ball, the physical shape of the fist closely matches the tennis ball's shape and size. However, some iconic gestures involve an imaginary component – a mental bridge between the gesture and the intended meaning (Werner & Kaplan, 1984). For example, when someone uses both hands to show the size of a tennis ball, the tennis ball becomes an imaginary component. To interpret iconic gestures in our study, infants had to mentally place the objects in between the hands of the actress. Even though the objects and gestures were displayed on the screen simultaneously, infants still had to imagine the object fitting in between the actress's hands. This may have made the task too challenging for 14-17-month-olds.

The Cross-Modal Correspondences

There are three ways in which the nature of our cross-modal stimuli could have led to null results. First, there may be other visual correlates to auditory pitch besides size that may have interfered in the current study. For example, spatial height may have interfered with the intended pitch-size cross-modal correspondence. Stumpf (1883) suggested that the cross-modal correspondence between pitch and size is based on the analogy between pitch and spatial height, where high pitch relates to high spatial locations and low pitch to low spatial locations. This cross-modal correspondence has also been shown in infants aged 3-4 months (Dolscheid et al., 2014). When we presented the visual objects to the infants, the large and the small version of the visual objects appeared at different heights on the screen. It is unlikely that the spatial height of the stimulus interfered with infants' sensitivity to the cross-modal correspondence between pitch and size, because the large visual object was always higher on the screen than the small visual object, and vice versa. If spatial height was driving the pitch-size cross-modal correspondence here, as suggested by Stumpf (1883), then we should have seen a significant effect, albeit in the opposite direction than predicted. Alternatively, it could be the case that both size and spatial height had an effect, but cancelled each other out, leading to null results. Future studies could distinguish between size and height in their stimulus presentation to further investigate this.

Second, our intended cross-modal correspondences may have been too complex for infants. The stimuli included a spoken non-word, a gesturing actress, and two objects. This multimodal setup, with cross-modal correspondences between speech, gesture, and object size, might have overwhelmed infants. Nevertheless, the lack of a multimodal effect aligns with previous research showing that 18-month-olds did not recognize iconic correspondences between a vocalization-gesture signal and a referent (Bohn et al., 2019).

Finally, there was a mismatch between the dynamicity of our cross-modal stimuli. While the objects on the screen remained static, the speech and gesture cues were more dynamic. Cross-modal correspondences are clearer when all stimuli share the same dynamic nature (either all static or all dynamic). Future research could explore fully dynamic cross-modal correspondences using objects that change in size, accompanied by matching auditory stimuli and gestures.

Conclusion

This study investigated whether 14-17-month-old infants use iconic speech and gesture cues to interpret words. We encoded object size (small or large) using speech pitches (high or low) and iconic gestures (small or large hand movements). However, neither cue elicited reliable preferences for congruently sized objects. For iconic speech cues, infants may not yet be sensitive to mappings between pitch and size at this age. Alternatively, our paradigm may not have elicited this cross-modal association. However, our null results do not preclude the possibility that a pitch-size correspondence does exist in infancy and may facilitate language development, as proposed by the sound symbolism bootstrapping hypothesis (Imai & Kita, 2014). For iconic gesture cues, infants may not yet understand iconic gestures at this age. Alternatively, our gestures may have been too complex. Finally, when infants were presented with iconic speech and gesture cues which indicated the same object size, they showed no reliable preference either. Nevertheless, our study does not rule out the possibility that there is a cross-domain iconicity effect in early language development.

Acknowledgements

This research was part of the undergraduate dissertation of LDS and funded by an Undergraduate Research Support Scheme (URSS) bursary from the University of Warwick awarded to LDS. We would like to thank the caregivers and infants for their voluntary participation in this study. We would also like to express our gratitude to Sevilay Şengül Yıldız who independently coded a third of the data as part of her Erasmus Research Visit to the University of Warwick.

References

Aussems, S., & Kita, S. (2019). Seeing iconic gestures with action events facilitates children's memory of these events. *Child Development, 90*, 1123-1137.

Aussems, S., & Kita, S. (2021). Seeing iconic gesture promotes first- and second-order verb generalization in preschoolers. *Child Development, 92*, 124-141.

Bohn, M., Call, J., & Tomasello, M. (2019). Natural Reference: A phylo- and ontogenetic perspective on the comprehension of iconic gestures and vocalizations. *Developmental Science, 22*, e12757.

Brown, R. (1958). *Words and things*. Glencoe, IL: Free Press.

de Saussure, F. (1983). *Course in general linguistics.* La Salle, IL: Open Court.

Dolscheid, S., Hunnius, S., Casasanto, D., & Majid, A. (2014). Prelinguistic infants are sensitive to space-pitch associations found across cultures. *Cognition, 132*, 334-341.

Faul, F., Erdfelder, E., Lang, A., & Buchner, A. (2007). G*Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior Research Methods, 39*, 175–191.

Fernández-Prieto, I., Navarra, J., & Pons, F. (2015). How big is this sound? Crossmodal association between pitch and size in infants. *Infant Behavior & Development, 38*, 77-81.

Fort, M., Lammertink, I., Peperkamp, S., Guevara-Rukoz, A., Fikkert, P., & Tsuji, S. (2018). Symbouki: A meta-analysis on the emergence of sound symbolism in early language acquisition. *Developmental Science, 21*, e12659.

Gallace, A., & Spence, C. (2006). Multisensory synesthetic interactions in the speeded classification of visual size. *Perception & Psychophysics, 68*, 1191-1203.

Herold, D., Nygaard, L., Chicos, K., & Namy, L. (2011). The developing role of prosody in novel word interpretation. *Journal of Experimental Child Psychology, 108*, 229-241.

Hodges, A., Özçalıskan, Ş., & Williamson, P. (2018). Type of iconicity influences children's comprehension of gesture. *Journal of Experimental Child Psychology, 166*, 327-339.

Imai, M., & Kita, S. (2014). The sound symbolism bootstrapping hypothesis for language acquisition and language evolution. *Philosophical Transactions of the Royal Society B: Biological Sciences, 369*, 20130298.

Köhler, W. (1947). *Gestalt psychology: An introduction to new concepts in modern psychology* (Rev. ed.). New York: Liveright Publishing.

Lausberg, H., & Sloetjes, H. (2009). Coding gestural behavior with the NEUROGES-ELAN system. *Behavior Research Methods, 41*, 841-849.

Lowe, M. L., & Haws, K. L. (2017). Sounds big: The effects of acoustic pitch on product perceptions. *Journal of Marketing Research, 54*, 331-346.

Maurer, D., Pathman, T., & Mondloch, C. (2006). The shape of boubas: Sound–shape correspondences in toddlers and adults. *Developmental Science, 9*, 316-322.

McNeill, D. (1992). *Hand and mind: What gestures reveal about thought.* Chicago, IL: University of Chicago Press.

Monaghan, P., Shillcock, R., Christiansen, M., & Kirby, S. (2014). How arbitrary is language? *Philosophical Transactions of the Royal Society B: Biological Sciences, 369*, 20130299.

Mondloch, C., & Maurer, D. (2004). Do small white balls squeak? Pitch-object correspondences in young children. *Cognitive, Affective, & Behavioral Neuroscience, 4*, 133-136.

Namy, L. (2008). Recognition of iconicity doesn't come for free. *Developmental Science, 11*, 841-846.

Nygaard, L., Herold, D., & Namy, L. (2009). The semantics of prosody: Acoustic and perceptual evidence of prosodic correlates to word meaning. *Cognitive Science, 33*, 127-146.

Ohala, J. (1984). An ethological perspective on common cross-language utilization of F0 of voice. *Phonetica, 41*, 1-16.

Ota, M., & Skarabela, B. (2016). Reduplicated words are easier to learn. *Language Learning and Development, 12*, 380-397.

Parise, C., & Spence, C. (2012). Audiovisual crossmodal correspondences and sound symbolism: A study using the implicit association test. *Experimental Brain Research, 220*, 319-333.

Perniss, P., & Vigliocco, G. (2014). The bridge of iconicity: From a world of experience to the experience of language. *Philosophical Transactions of the Royal Society B: Biological Sciences, 369*, 20130300.

Perry, L., Perlman, M., Winter, B., Massaro, D., & Lupyan, G. (2018). Iconicity in the speech of children and adults. *Developmental Science, 21*, e12572.

Reinisch, E., Jesse, A., & Nygaard, L. (2013). Tone of voice guides word learning in informative referential contexts. *Quarterly Journal of Experimental Psychology, 66*, 1227-1240.

Shintel, H., Nusbaum, H., & Okrent, A. (2006). Analog acoustic expression in speech communication. *Journal of Memory & Language, 55*, 167-177.

Stumpf, C. (1883). *Tonspsychologie* [Tone Psychology].

Tomasello, M., Striano, T., & Rochat, P. (1999). Do young children use objects as symbols? *British Journal of Developmental Psychology, 17*, 563-584.

Werner, H., & Kaplan, B. (1984). *Symbol Formation* (1st ed.). Psychology Press.