The facilitatory role of sound symbolism in infant word learning

Michiko Miyazaki[1] (myzk@lab.tamagawa.ac.jp), Shohei Hidaka[2] (shhidaka@jaist.ac.jp), Mutsumi Imai[3] (imai@sfc.keio.ac.jp), H. Henny Yeung[4] (henny.yeung@parisdescartes.fr), Katerina Kantartzis[5] (k.f.kantartzis@bham.ac.uk), Hiroyuki Okada[1] (h.okada@eng.tamagawa.ac.jp), Sotaro Kita[5] (S.Kita@bham.ac.uk)

[1]Brain Science Institute, Tamagawa University, JAPAN,
[2]School of Knowledge Science, Japan Advanced Institute of Science and Technology, JAPAN
[3]Department of Environment and Information Studies, Keio University, JAPAN
[4]Université Paris Descartes, & CNRS (Laboratoire Psychologie de la Perception, UMR 8158), FRANCE
[5]School of Psychology, University of Birmingham, UK

Abstract

Sound symbolism or the nonarbitrary link between language sound and meaning are commonly found across many languages of the world. A well-known example is the association between rounded vs. angular shapes and labels (i.e., the Bouba-kiki effect by Köhler, 1929/1947). Previous research has shown that sound symbolic words play facilitative role for preschool children's novel verb learning (Imai, Kita, Nagumo & Okada, 2008; Kantartiz, Imai & Kita, 2011), helping children identify what aspects of motion events should be mapped to verbs. In this research, we explore whether sound symbolism may facilitate language learning in human infants who have just begun to learn word meanings. Sound symbolism may be a useful cue particularly at the earliest stages of word learning, because this cue seems to be available without needing prior word learning experience (Gogate & Hollich, 2010). Using a habituation paradigm and a Bayesian model-based analysis, we demonstrated that 14-month-old infants could detect Köhler-type (1947) shape-sound symbolism, and could use this sensitivity in their effort to establish the word-referent association.

Keywords: Sound symbolism; Word learning; Iconicity of language; Origin of language; Multisensory mapping; Bayesian analysis

Introduction

Traditional linguistics has long assumed that links between a word's form and meaning are arbitrary (de Saussure, 1916/1983). However, words whose forms are motivated by their meanings (i.e., sound symbolic words) are commonly found across many languages of the world. For example, *bump* and *thump* sound like what they mean: an event with an abrupt end (e.g., Firth, 1935). Several languages even have large grammatically defined lexical classes of sound symbolic words (i.e., "ideophones," "expressives," or "mimetics") (Voeltz & Kilian-Hatz, 2001; Kita, 1997). A well-known example of sound symbolism is the association between rounded vs. angular shapes and labels (Köhler, 1929/1947; see also Ramachandran & Hubbard, 2001). Given a forced choice, adults and children from different languages (e.g., German, English, and Swahili) much prefer to label rounded objects *bouba* (or *maluma)* and angular objects *kiki* (or *takete)* (e.g., Davis, 1961; Holland & Wertheimer, 1964).

Recently, evidence for sensitivity to sound symbolism in young infants is emerging (Penã, Mehler, & Nespor, 2011; Ozturk, Krehm, & Vouloumanos, 2012). Ozturk et al. (2012) demonstrated that even 4-month-olds are sensitive to the bouba-kiki sound-shape mappings.

An interesting question is whether sensitivity to sound symbolism is useful for language development. Previous research has shown that sound symbolic words play facilitative role for preschool children's novel word learning, helping children identify what aspects of the world should be mapped to verbs (Quine, 1969, Markman, 1989). For example, Maurer et al. (2006) demonstrated that 2.5-year-old children are likely to match rounded sounding labels with simple, rounded objects and jagged sounding labels with pointy objects (Mondloch & Maurer, 2004; Maurer, Pathman, & Mondloch, 2006). In addition, sound symbolism also helps 3-year-olds infer meanings of novel verbs, which are generally more difficult than object names (Imai et al., 2008; Kantartzis, Imai, & Kita, 2011). Thus, sound symbolism is a good candidate for mappings from word forms to referents in the early stages of development. However, it is not clear whether such sensitivity to sound symbolism in young infants is used for word learning at the initial stage of language development.

Before toddlerhood, it is unknown how the youngest word learners can "break into" the process of mapping words to their referents. For infants just starting to learn words, the induction problem is much harder (Hollich, et al., 2007; Spiegel & Halberda, 2011), and these infants likely rely more on perceptual regularities, due to limited memory and information processing abilities (Gogate & Hollich, 2010). In fact, infants between 10- and 17-months of age find it much more difficult to utilize word-learning strategies used by older infants to infer novel word meanings (Woodward & Hoyne, 1999). At this age, infants may use more perceptual strategies to map words, which can often lead them to the wrong referent (e.g., Hollich et al., 2007).

As with toddlers, however, sound symbolic words may be easier to learn, and these early links may help to scaffold mappings from word forms to referents in the early stages of development. In the present study, we examine whether sound symbolic links can provide early word-referent cues. Specifically, we asked if 14 month-old infants could utilize senseitivity to Köhler-type (1947) shape-sound symbolism to establish association between a word and a referent object. Here we chose infants at 14 months of age, who are just old enough to learn new words in laboratory tasks, but are at an age when learning is still very fragile, especially when trying to process the precise phonological forms of words (Stager & Werker, 1997).

We hypothesized that 14-month-olds detect sound-shape correspondence, and that this ability helps these infants make mappings from word forms to their referents. Specifically, we taught Japanese-learning infants two word labels, and then tested whether they encoded these labels in a preferential looking procedure. Half the infants were taught two sound symbolic labels (as rated by adults); half were taught the mismatching labels. We predicted that those in the sound symbolically matching condition would learn labels more easily than those in the mismatching condition.

This report features a Bayesian model-based data analysis. Infants looking behavior—particularly in a preferential looking paradigm, in which infants must compare two objects to make a decision—is very complex: It often shifts dynamically instead of staying stable during a trial. Also, because of the dynamic nature of looking, it is difficult to determine the most appropriate time window prior to the analysis. Furthermore, infants' looking time is "noisy", affected largely by looking profiles inherent in individual infants (e.g., a preference to look at a certain location on the monitor, a preference to look at a particular object, the likelihood to shift eye-gaze, etc.). Ignoring these idiosyncratic looking biases can weaken statistical power. Despite this problem, researchers have traditionally used ANOVA to examine the effect, where infants' looking time is averaged over a pre-set time window and individual-based looking biases are treated as experimental noise. In this research, we employed a Bayesian approach to analyze the data, which offers a new method of data analysis that is more adaptive to complex and dynamic nature of infants' looking behavior (see the Bayesian Data Analysis section for more details).

Method

Participants

Participants were thirty-four full term, monolingual Japanese 14-month-olds (M = 14;16, range 13;27-15;9, 22 males). Infants were randomly assigned to either the match or mismatch condition. An additional 11 infants were excluded from data analyses due to experimental error (n = 1), or fussiness during the experiment (n = 10).

Apparatus

A black cloth curtain surrounded a 21-inch display where visual stimuli appeared, and a digital video camera was hidden below the screen, relaying video of the infant to the control area. Video was also recorded (29.97 fps) for offline coding of looking.

Stimulus materials

Target stimuli consisted of a round versus a spiky shape, as well as audio recordings of two novel words, *kipi* and *moma* (Figure 1). Stimuli were constructed on the basis of a pilot study, where we first chose consonants (m, l, n, p, k), and vowels (a, o, i) that are related to smooth and jagged shapes according to previous research on sound symbolism (Köhler, 1929/1947; Maurer, et al., 2006; Westbury, 2005). *Moma* and *kipi* were selected because our adults frequently chose these as the best match to the smooth/round and spiky/jagged shapes, respectively, and both were nonwords in Japanese. The two shapes were colored to make them more interesting to infants. Colors were chosen not to affect the shape sound symbolism.

During a habituation phase, infants were presented with two pairs of audio-visual stimuli: in the match condition, *kipi* – spiky, object and *moma* - round object; in the mismatch condition, *kipi* - round object, *moma* - spiky object. Infants were presented with filler stimuli, consisting of colored drawings of a ball, a banana, a car, and a picture book, paired with audio recordings of the corresponding word in Japanese. These items were chosen based on normative data from the Japanese MacArthur-Bates Communicative Development Inventory (J-MCDI) (Ogura & Watamaki, 2004). A female Japanese speaker recorded the target and filler words, along with all other speech used in the experiment, in an infant-directed speech register.

Procedure

Infants were tested individually in a quiet room on their parent's lap, positioned 60 cm from the display. Pretest, habituation, and test phases were presented, each of which contained several trials separated by a short attention-getting movie. Parents were instructed to keep their eyes closed, and also asked to complete the J-MCDI.

Pretest phase Here we presented 4 familiarization and 2 referential trials in random order. Familiarization trials showed side-by-side displays of either filler or target objects. Each trial lasted 8 seconds and was accompanied by, "Mite! Mite!" (Look! Look!). Two referential trials were included to enhance infants' understanding of the referential nature of the labels (Fennell & Waxman, 2010), and here a single familiar object slowly moved (either up and down, or right and left) on the display, accompanied by the corresponding label in isolation and then in a carrier sentence (e.g., "Kuruma! Kuruma-wo mite!" [Car! Look at the car!]).

Habituation phase The habituation phase consisted of a pseudo-randomly ordered series of trials such that each word-object pairing appeared twice in every block of four trials. In each trial, a single target object moved slowly from right to left in order to maintain infants' attention (Werker et al., 1998), while one of the target words was paired with it (Figure 1). Trials lasted a maximum of 16 seconds and were accompanied by 13 tokens of a target word, each spoken with a different intonation pattern. The habituation criterion was set to a maximum of 24 trials, or a decrease of 65% in looking to the longest previous block of 4 trials (Stager &

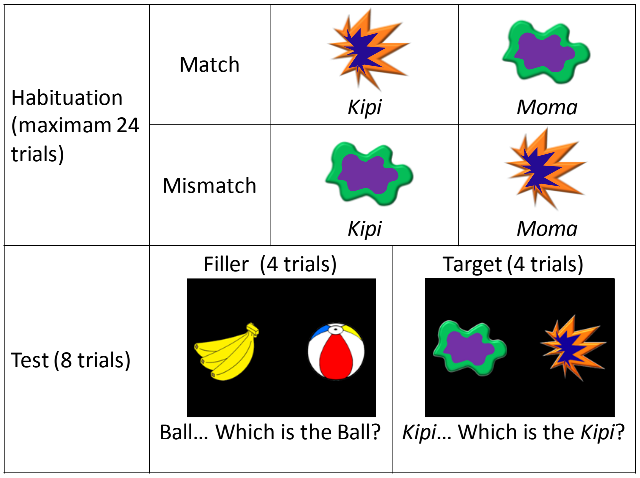

Figure 1: Design and Stimuli

Werker, 1997; Yoshida, Fennell, Swingley, & Werker, 2009).

Test phase The test phase consisted of four filler and four target trials that contained two objects side-by-side, counterbalanced for side over the testing phase. Two filler trials were always shown first, providing infants the opportunity to see familiar objects in the test procedure. The third and fourth trials were target trials, and the remaining four trials alternated between filler (F) and target (T) trials. All trials lasted 8 seconds, and began with a 3000 ms silent baseline period to measure visual preferences for the objects, followed by a phrase asking infants to look at the correct object (i.e., "X! X wa docchi?" [X! Which is the X?], where X stands for a filler or target word).

Coding

Two blind coders classified each video frame as a look to the right, to the left, an ambiguous look, or no look. Fifteen percent of the all samples were coded twice to obtain inter-coder reliability. The inter-coder reliability was high, $\kappa = .826$.

Bayesian data analysis

Data were analyzed using a Bayesian approach. In analyzing infants' looking behavior, it is customarily assumed that it takes a certain time for infants to process auditory stimuli and/or move their eyes to fixate the target stimulus (e.g., Swingley & Aslin, 2002). Considering the lag due to auditory and visual processing, a pre-determined time window is set, and infants' looking is averaged throughout this window. The group means are then compared using a linear model (typically a t-test or ANOVA). However, as mentioned earlier, the assumptions in traditional linear models often pose limitations for analyzing infants' complex looking behavior. First, infants' looking to the target often shifts dynamically instead of staying stable during the time course, and hence polynomial functions appear to fit better than a linear function. Second, because infants' looking behavior may be affected by the nature of the stimuli and experimental settings, it is difficult to determine the most appropriate time window prior to the analysis. Third, although substantial individual difference is expected in infants' looking profiles (e.g., a preference to look at a certain location on the monitor, a preference to look at a particular object, the likelihood to shift eye-gaze, etc.), these response biases are simply treated as experimental noise, which weakens statistical power to detect the experimental effect of interest.

In the current analysis—to rectify these limitations in traditional linear models—we performed a Bayesian model-based analysis based on Yurovsky, Hidaka, and Wu, 2012. In a Bayesian framework, a hypothetical relationship among a set of experimental factors (e.g., training effects), subsidiary factors (e.g., object-specific bias) and potential patterns of behaviors (e.g., looking time) are expressed as a statistical model with a set of parameters which is estimated through model fitting to a given dataset. In the present analysis, there are two major sets of parameters. The first set includes preference parameters for factors that could potentially affect the looking of a particular AOI at a given moment: individual infants' location preference, object preference, preference to look at the trained object, preference to look at the object that was sound symbolically matching to the label. The second set consisted of group parameters, which classify participants in such a way that within-group similarity in infants' response patterns and across-group differences are simultaneously maximized.

Results

Looking data from each of the critical video frames were classified as a look to the left AOI, to the right AOI, or to a 'no-look' AOI. We analyzed infants' looking times (as frame-by-frame counts) as a function of five factors: a location-specific preference, an object-specific preference, a "correct" (trained) object preference, a sound symbolism preference, and the interaction between training and the sound symbolic match. The location-specific preference is defined as a bias to look toward the left or right AOI, relative to the preference for the no-look AOI for each infant independent of the match/mismatch condition. The object preference is a bias to look toward a particular object compared to the no-look AOI (with some other objects). The correct object preference is a preference to look at the object with which the label was associated during the habituation phase. The sound symbolism preference is a tendency to look at the sound symbolically matching object (to the label heard in the trial) after the onset of the speech (i.e., the preference for the spiky object after the onset of speech "kipi," or the preference for the round object after "moma"). Through the process of model fitting, we estimated a set of parameters for all of the five factors above, but we focus only on the three experiment-relevant factors, i.e., training, sound-symbolic match, and training-sound-symbolism interaction factors here.

We first analyzed the effect of the training and that of sound symbolism separately against the baseline preference, to see whether training and/or sound symbolism alone affected infants' looking behavior after the speech onset. For this purpose, we performed a series of models using (Bayesian) hypothesis testing. In Bayesian hypothesis testing, each model specifies a probability of a hypothesis to reproduce the current looking dataset. We tested a contribution of a particular factor, by evaluating the goodness of fit for two models—one with the target factor and one without—by a *Bayes factor* (Jeffreys, 1961; See also Wagenmakers et al. (2010) for a review in psychological studies). "The Bayes factor (BF) *X* of Model *A* to *B* given a dataset" indicates that the odds ratio for Model *A* to reproduce the dataset is *X* times higher than that for Model *B* to do so under even prior probability for each model[1]. In the present study, we considered a Bayes factor larger than 30 (or equivalently log-Bayes factor larger than 3.4) to be strong evidence in support for the Model A over Model B, based on Jeffreys' (1961) criterion.

Six models were evaluated as shown in Table 1. The first three models—P1-full, P2-full, and P3-full models--were full models in which all of the three experimental factors were included. These three models assumed different levels of complexity in their functions of looking probability. The first order polynomial function was assumed for the P1-full model, and the second- and third-order polynomial functions were assumed for the P2-full and P3-full models, respectively. We then compared the three models to determine the optimal level of complexity of the function in the model. The analysis on the Bayes factors indicates that the middle degree of complexity (P2-full) is strongly favored over both the relatively simple (P2-full to P1-full: 27.5) and complex model (P2-full to P3-full: 247.4). We therefore employed the P2-full model as the baseline model, against which each of the subset models was compared.

To evaluate the effect of the three experimental factors, we calculated the Bayes factors for the three additional models in Table 1, i.e., P2-NoInt., P2-NoSS, and P2-NoTr against the P2-full baseline model (see Table 1 for results and the abbreviations). The analysis of the Bayes factors suggests that the P2-full model was strongly favored over the P2-NoInt (109.6), P2-NoSS (119.6), and the P2-NoTr (92.5). These results indicate that all effects of sound symbolism, training and the interaction between training and sound symbolism significantly contributed to the model fit.

This suggest that infants tended to look at the trained object regardless of whether this object was sound symbolically matching or not. Furthermore, sound symbolism affected infants' looking, regardless of whether infants were trained on a sound symbolically matching object.

[1] According to Jeffreys (1961), BF from 3 to 10 (log-BF from 1.1 to 2.3) indicates "substantial" evidence, BF from 10 to 30 (log-BF from 2.3 to 3.4) indicates "strong" one, and BF from 30 (log-BF greater than 3.4) indicates "very strong" one.

Table 1: Summary of the hypotheses testing on the six models.

	Models					Results	
Models	#P	Tr	SS	TS	Ply	#G	log-BF
P2-full*	10	Y	Y	Y	2	7	0
P1-full	7	Y	Y	Y	1	7	27.5
P3-full	13	Y	Y	Y	3	6	247.4
P2-NoInt.	8	Y	Y	-	2	7	109.6
P2-NoSS.	8	Y	-	Y	2	7	119.6
P2-NoTr.	8	-	Y	Y	2	8	92.5

The abbreviations are as follows: #P: the number of parameters for each cluster, Tr, SS, and TS: whether the model contains preference parameters for trained object, sound symbolism match object, and the interaction between them ("Y" if the model has), Ply: the order of polynomial functions of the looking time courses , #G: the estimated number of groups of infants, Log-BF: the log-Bayes factor of the P2-full model reproducing the data relative to the compared model. P*n*-full: a full-factored model with the *n*-th order polynomial functions, P2-NoInt: a model without interaction between training and sound symbolic match, P2-NoSS: a model without the effects of sound-symbolic match, P2-NoTr: a model without the effects of training, *: the best model.

Discussion

The present study aimed to clarify the facilitating role of sound symbolism in novel word learning in 14-month-old infants. Although the infants at this age sometimes show difficulty for matching words to their correct referents due to their limited cognitive ability, 14-month-olds could utilize sound-symbolic correspondences between speech sounds and object in this study. By Bayesian analysis, the effects of sound symbolism, training, and the interaction between sound symbolism and training all significantly contributed to the infants' looking behavior.

The current Bayesian data analysis shed light on the issue of how the youngest word learners break into the incredibly difficult process of mapping words to their references. One of the great advantages of Bayesian analysis is that it could wash out the critical experimental effects when traditional averaged analysis are treated as "noise," can be considered. By classifying participants with similar looking patterns into clusters and estimating group parameters, fine-grained response characteristics for a particular subgroup of infants.

The fact that significant contribution of factors sound symbolism, training, and the interaction between sound symbolism and training revealed in early word leaning stage may provide an important clue for solving the difficulty for mapping word to their referents. Sound symbolism may allow infants to anchor speech to meaning, which in turn helps them obtain "referential insight"--- the insight that language sounds are symbols that represent concepts (Gogate & Hollich, 2010). Once infants get into sound-symbolically based systems relating surface structure to meaning, they may be able to use this early knowledge to bootstrap themselves to more abstract meanings, needing direct perceptual anchors less and less, perhaps reflecting similar trajectories in language evolution (Kita, Kantartzis, & Imai, 2010).

Acknowledgments

We would like to thank Janet Werker for insightful discussion and advice in designing the experiment. Mamiko Arata, Noburo Saji and Junko Kanero all contributed to the creation and selection of stimuli. Asako Isawa, Kaoru Takata and Satomi Kanbara contributed to the video coding. This research was supported by MEXT KAKENHI (#15300088, #22243043, #23120003) for Imai, Okada, and (#23300099) for Hidaka, MEXT GCOE program to Tamagawa University, and Biotechnology and Biological Sciences Research Council's Research Development Fellowship [BB/G023069/1] to Kita.

References

Davis, R. (1961). The fitness of names to drawings. A cross-cultural study in Tanganyika. *British Journal of Developmental Psychology, 52*, 259-268.

de Saussure, F. (1916/1983). *Course in general linguistics* (R. Harris, Trans.). La Salle, IL: Open Court.

Fennell, C. T., & Waxman, S. R. (2010). What paradox? Referential cues allow for infant use of phonetic detail in word learning. *Child development, 81*(5), 1376-1383.

Firth, J. R. (1935). The technique of semantics. *Transactions of the Philological Society*, 36-72.

Gogate, L. J., & Hollich, G. (2010). Invariance detection within an interactive system: a perceptual gateway to language development. *Psychological review,117*(2), 496-516.

Holland, M. K., & Wertheimer, M. (1964). Some physiognomic aspects of naming, or maluma and takete revisited. *Perceptual and Motor Skills, 19*(1), 111-117.

Hollich, G., Golinkoff, R. M., & Hirsh-Pasek, K. (2007). Young children associate novel words with complex objects rather than salient parts. *Dev Psychol,43*(5), 1051-1061.

Imai, M., Kita, S., Nagumo, M., & Okada, H. (2008). Sound symbolism facilitates early verb learning. *Cognition, 109*(1), 54-65.

Jeffreys, H. (1961). *The Theory of Probability* (3rd ed.). Oxford.

Köhler, W. (1929/1947). *Gestalt psychology: An introduction to new concepts in modern psychology*. New York: Liveright Publishing Corporation.

Kantartzis, K., Imai, M., & Kita, S. (2011). Japanese Sound-Symbolism Facilitates Word Learning in English-Speaking Children. *Cognitive Science, 35*, 575-586.

Kita, S. (1997). Two-dimensional semantic analysis of Japanese mimetics. *Linguistics, 35*, 379-415.

Kita, S., Kantartzis, K., & Imai, M. (2010). Children learn sound symbolic words better: Evolutionary vestige of sound symbolic protolanguage. *Paper presented at the The evolution of language: Proceedings of the 8th International Conference (EVOLANG8)*, 206-213. Singapore: World Scientific.

Markman, E. M. (1989). *Categorization and naming in children: Problems of induction*. Cambridge, Mass.: MIT Press.

Maurer, D., Pathman, T., & Mondloch, C. J. (2006). The shape of boubas: sound-shape correspondences in toddlers and adults. *Developmental science, 9*(3), 316-322.

Mondloch, C. J., & Maurer, D. (2004). Do small white balls squeak? Pitch-object correspondences in young children. *Cognitive, affective & behavioral neuroscience, 4*(2), 133-136.

Ogura, T., & Watamaki, T. (2004). *MacArthur Communicative Development Inventories: User's Guide and Technical Manual.* Kyoto: Kyoto International Social Welfare Exchange Centre.

Ozturk, O., Krehm, M. & Vouloumanos, A. (2013). Sound symbolism in infancy: evidence for sound–shape cross-modal correspondences in 4-month-olds. Journal of Experimental Child Psychology, 114, 173-186.

Pena, M., Mehler, J., & Nespor, M. (2011). The role of audiovisual processing in early conceptual development. *Psychological science, 22*(11), 1419-1421.

Quine, W. V. (1960). *Word and Object:* The MIT Press.

Ramachandran, V. S., & Hubbard, E. M. (2001). Psychophysical investigations into the neural basis of synaesthesia. *Proc Biol Sci, 268*(1470), 979-983.

Spiegel, C., & Halberda, J. (2011). Rapid fast-mapping abilities in 2-year-olds. *Journal of experimental child psychology, 109*(1), 132-140.

Stager, C. L., & Werker, J. F. (1997). Infants listen for more phonetic detail in speech perception than in word-learning tasks. *Nature, 388*(6640), 381-382.

Swingley, D., & Aslin, R. N. (2002). Lexical neighborhoods and the word-form representations of 14-month-olds. *Psychological science, 13*(5), 480-484.

Voeltz, F. K. E., & Kilian-Hatz, C. (Eds.). (2001). *Ideophones*. Amsterdam/Philadelphia: John Benjamins.

Wagenmakers, E-J., Lodewyckx, T., Kuriyal, H., & Grasman, R. (2010). Bayesian hypothesis testing for psychologists: A tutorial on the Savage-Dickey method., *Cognitive Psychology*, 60, 158-189.

Werker, J. F., Cohen, L. B., Lloyd, V. L., Casasola, M., & Stager, C. L. (1998). Acquisition of word-object associations by 14-month-old infants. *Dev Psychol, 34*(6), 1289-1309.

Westbury, C. (2005). Implicit sound symbolism in lexical access: evidence from an interference task. *Brain and language, 93*(1), 10-19.

Woodward, A. L., & Hoyne, K. L. (1999). Infants' learning about words and sounds in relation to objects. *Child development*, *70*(1), 65-77.

Yoshida, K. A., Fennell, C. T., Swingley, D., & Werker, J. F. (2009). Fourteen-month-old infants learn similar-sounding words. *Developmental science*, *12*(3), 412-418.

Yurovsky D, Hidaka S, & Wu R (2012) Quantitative Linking Hypotheses for Infant Eye Movements. PLoS ONE 7(10).