The Role of Audiovisual Processing in Early Conceptual Development

Marcela Peña, Jacques Mehler and Marina Nespor

Psychological Science published online 29 September 2011
DOI: 10.1177/0956797611421791

Psychological Science OnlineFirst, published on September 29, 2011 as doi:10.1177/0956797611421791

Short Report

Psychological Science
XX(X) 1–3

DOI: 10.1177/0956797611421791
http://pss.sagepub.com

The Role of Audiovisual Processing in Early Conceptual Development

Marcela Peña[1,2], Jacques Mehler[2], and Marina Nespor[2,3]
[1]Laboratorio de Neurociencias Cognitivas, Escuela de Psicología, Pontificia Universidad Católica de Chile; [2]Language, Cognition and Development Lab, International School for Advanced Studies (SISSA), Trieste, Italy; and [3]Department of Psychology, University of Milano-Bicocca

Received 3/28/11; Revision accepted 6/22/11

Historically, the link between the sounds and meanings of words has been reported as being arbitrary (de Saussure, 1959). Some observations, however, support the idea that word-meaning associations rely on systematic relations between the physical properties of speech sounds and their referents (Brown, 1958; Ohala, 1994; Sapir, 1929).

In pioneering studies, Edward Sapir (e.g., Sapir, 1929) invented pairs of pseudowords, such as *mal* and *mil*, and asked adults to match them with the notions of "big" and "small." Participants preferred using words with the sound [a] to represent the idea of "big" and words with the sound [i] to represent the idea of "small." Indeed, across several languages, mostly high frontal vowels (e.g., [i] and [e]) are used in diminutive morphemes, and mostly low posterior vowels (e.g., [o] and [a]) are used in augmentative morphemes (Ohala, 1994).

Young children are also sensitive to nonarbitrary links between speech and visual stimuli. For instance, 3-year-old children prefer to use pseudowords such as *maluma* and *bouba* (containing posterior vowels) to label objects with blunt angles and pseudowords such as *kiki* (containing frontal vowels) to label objects with sharp angles (Maurer, Pathman, & Mondloch, 2006). Moreover, 3-year-old children can exploit the acoustic-phonetic regularities of foreign words to learn word-meaning associations. Indeed, learning of pseudoverbs referring to walking-related actions is facilitated among English-speaking children if the novel verbs sound like real Japanese walking-related verbs, and is facilitated among Japanese-speaking children if the pseudoverbs sound like English walking-related verbs (Imai, Kita, Nagumo, & Okada, 2008; Kantartzis, Kita, & Imai, 2011).

We explored possible links between sounds and visual stimuli in young infants. Neonates are sensitive to acoustic-phonetic cues in language (Shi, Werker, & Morgan, 1999) and have notable cross-modal capacities (Lewkowicz & Turkewitz, 1980). Cross-modal capacities might contribute to conceptual development through their use in learning the relationships among a group of multimodal exemplars. We evaluated the preferences of 4-month-old infants for various pairings of sounds and visual objects. The sounds were monosyllables containing [i] or [o] (Experiment 1) or [e] or [a] (Experiment 2). The visual objects were shapes that differed in size. Each monosyllable was paired with two shapes, one large and one small. Infants' preferences were estimated using eye tracking.

Method

Participants were healthy infants from monolingual Spanish environments: 28 in Experiment 1 and 28 in Experiment 2 (mean age = 4 months, *SD* = 18 days; Experiment 1: 16 females and 12 males; Experiment 2: 20 females and 8 males). We excluded 4 infants from Experiment 1 and 6 infants from Experiment 2 because they did not complete the experimental protocol.

Infants were held in the lap of a parent and tested in a soundproof room. The parents wore a mask that prevented them from seeing the visual stimuli. In each of 32 trials, the infants simultaneously heard a syllable and watched an image presented on an eye-tracker screen (Tobii 1750, Stockholm, Sweden) that recorded the infants' eye movements.

Syllables were synthesized and contained a vowel ([i], [e], [a] or [o]) preceded by a consonant ([l], [f] or [d]). Syllables lasted 430 ms and were repeated five times per trial at 1,000-ms intervals through a loudspeaker situated centrally behind the eye-tracker screen. The fundamental frequency of the syllables containing [i] and [e] was 210 Hz, and the fundamental frequency of the syllables containing [o] and [a] was 190 Hz; these frequencies simulate those observed in natural languages (Ohala, 1997). Auditory stimuli containing frontal or posterior vowels were presented in a random order. In Experiment 1, auditory stimuli consisted of syllables with [i] or [o], whereas in Experiment 2, auditory stimuli were syllables that contained [e] or [a].

Corresponding Author:
Marcela Peña, SISSA, Neuroscience Sector, Via Bonomea, 265, 34136, Trieste, Italy
E-mail: pena@sissa.it

The visual stimuli in each trial consisted of a large exemplar and a small exemplar of an object; the exemplars were presented simultaneously, one on the right side of the screen and one on the left side, with their inferior edges aligned along the same horizontal axis. The large and small exemplars were identical except in size; each was presented on the left side half of the time and on the right side half of the time. The objects were the outlines of geometric shapes (circle, oval, square, or triangle) drawn in one of four possible colors (red, green, blue, or yellow). The outlines of the large exemplars reached the extremes of the 3.6- × 3.6-cm screen, whereas the small exemplars covered 72% of the screen. The order in which pairs were presented was random and counterbalanced across participants.

Results

Figure 1 illustrates the infants' viewing preferences as a function of the vowel in the monosyllable presented. Preferences were estimated by analyzing the direction of the first gaze (i.e., toward the larger object or the smaller object) and the total looking time toward each object (large vs. small). For each vowel sound, direction of the first gaze was calculated as the mean percentage of trials in which the first gaze was toward the larger object and the mean percentage of trials in which the first gaze was toward the smaller object. For each experiment, the means of these variables were submitted to two separate repeated measures analyses of variance with vowel ([i] vs. [o] in Experiment 1; [e] vs. [a] in Experiment 2) and object size

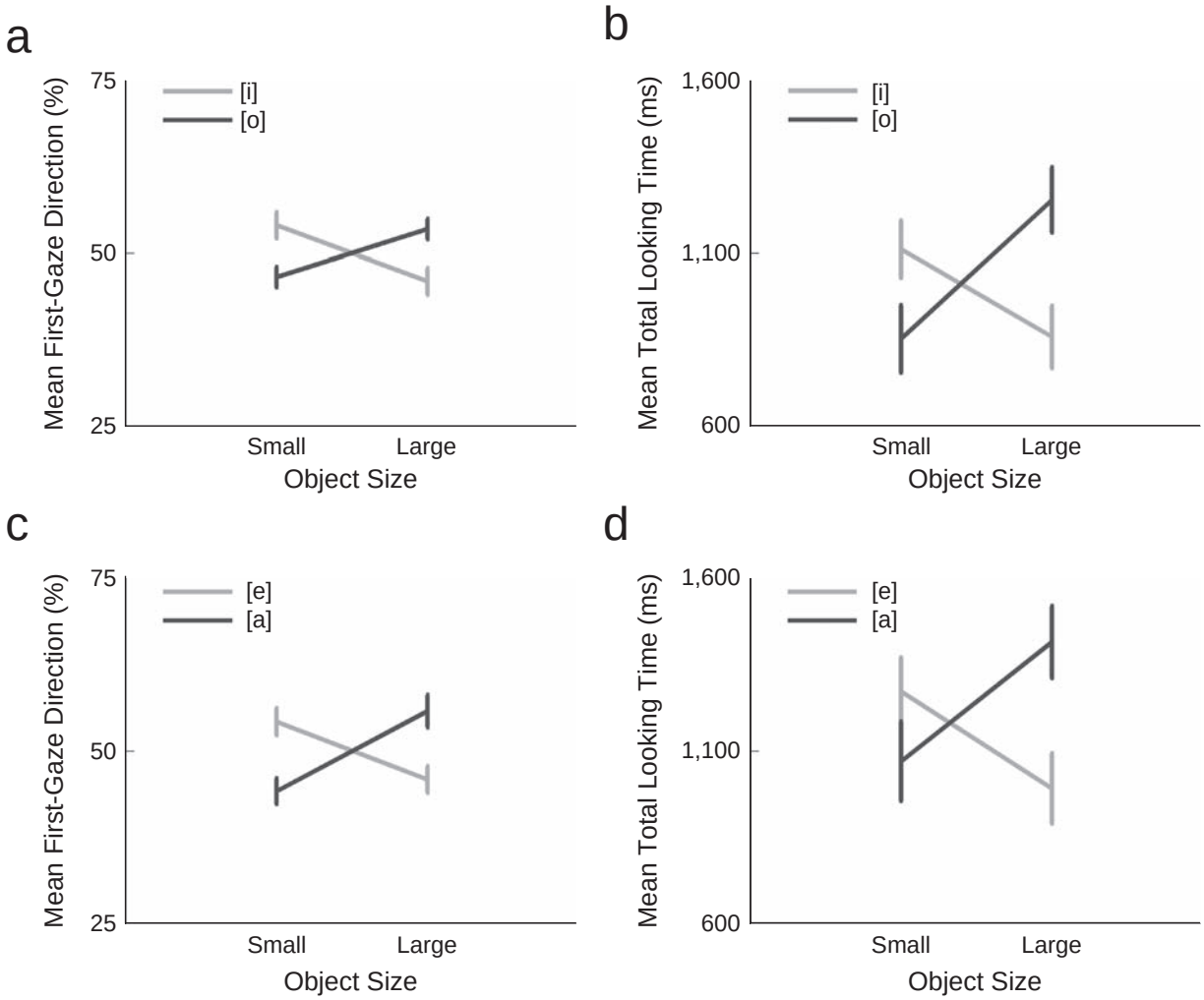

Fig. 1. Direction of the first gaze and total looking time in (a, b) Experiment 1 and (c, d) Experiment 2. The graphs in (a) and (c) show the percentage of trials in which the first gaze was toward the small object and the percentage of trials in which the first gaze was toward the large object as a function of the vowel in the syllable presented simultaneously with the objects. The graphs in (b) and (d) show the total looking times for the small object and the large object as a function of the vowel in the syllable presented simultaneously with the objects. The endpoints of the graphed lines correspond to mean values; error bars indicate standard errors of the mean.

(large vs. small) as within-subjects variables. The Greenhouse-Geisser correction for repeated measures was applied.

The analysis of first-gaze direction revealed a significant Vowel × Object Size interaction in both experiments—Experiment 1: $F(1, 27) = 13.002$, $p < .001$, $\eta^2 = .325$; Experiment 2: $F(1, 23) = 10.237$, $p < .004$, $\eta^2 = .308$. In other words, infants directed their first gaze significantly more often to large objects than to small objects when the syllables contained [o] or [a] and directed their first gaze significantly more often to small objects than to large objects when the syllables contained [i] or [e].

The analysis of total looking time also revealed a significant Vowel × Object Size interaction in both experiments—Experiment 1: $F(1, 27) = 13.188$, $p < .001$, $\eta^2 = .328$; Experiment 2: $F(1, 23) = 7.9$, $p < .010$, $\eta^2 = .256$. Infants spent more time looking at large objects than at small objects when they heard syllables containing [o] or [a], and they spent more time looking at small objects than at large objects when they heard syllables containing [i] or [e].

Discussion

Our results suggest that, like adults, 4-month-old infants can integrate what they hear and see into a single notion of size. High frontal vowels were significantly associated with small object size, and low posterior vowels were significantly associated with large object size. Infants' early audiovisual associations might reflect prior knowledge (innate or based on experience) of their environment, as postulated by Bayesian learning (Shultz, 2007), but it could also reflect amodal representations of the notion of size, such as those suggested by phonetic symbolism (Brown, Black, & Horowitz, 1955; Hinton, Nichols, & Ohala, 1994). Young infants have outstanding cross-modal and representational capacities (for a review, see Lewkowicz & Ghazanfar, 2009), such as those observed in neonates when they imitate facial expressions just hours after birth (Meltzoff, 1990). Previous experience in seeing the size of the mouth opening when humans utter different vowels may have prompted the infants in our study to match different vowels to different sizes, a phenomenon that might be at the root of the infants' notion of size. Further studies with congenitally blind infants might help to explore this hypothesis.

To the best of our knowledge, our results are the first to suggest that 4-month-old infants are able to infer size-related notions from an integrated analysis of speech sounds and visual objects. This finding suggests that a primitive capacity for pairing sensory stimulation from different modalities influences some aspects of conceptual development.

Acknowledgments

We are grateful to Enrica Pittaluga, Margarita Luna, Elizabeth Pavez, Francesca Gandolfo, and Marijana Sjekloca for their help recruiting infants. We also thank the infants and parents who participated in this study.

Declaration of Conflicting Interests

The authors declared that they had no conflicts of interest with respect to their authorship or the publication of this article.

Funding

This work was supported by James S. McDonnell Foundation Grant 21002089 to J. M. and by the following grants from the Comisión Nacional de Investigación Científica y Tecnológica de Chile to M. P.: Fondecyt 1090662, PIA-Conicyt-CIE-05, and PBCT-PSD72.

References

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

Brown, R. W., Black, A. H., & Horowitz, A. E. (1955). Phonetic symbolism in natural languages. *Journal of Abnormal and Social Psychology*, *50*, 388–393.

de Saussure, F. D. (1959). *Course in general linguistics*. New York, NY: Philosophical Library.

Hinton, L., Nichols, J., & Ohala, J. J. (1994). *Sound symbolism*. Cambridge, England: Cambridge University Press.

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

Kantartzis, K., Imai, M., & Kita, S. (2011). Japanese sound-symbolism facilitates word learning in English-speaking children. *Cognitive Science*, *35*, 575–586.

Lewkowicz, D. J., & Ghazanfar, A. A. (2009). The emergence of multisensory systems through perceptual narrowing. *Trends in Cognitive Sciences*, *13*, 470–478.

Lewkowicz, D. J., & Turkewitz, G. (1980). Cross-modal equivalence in early infancy: Auditory-visual intensity matching. *Developmental Psychology*, *16*, 597–607.

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

Meltzoff, A. N. (1990). Towards a developmental cognitive science: The implications of cross-modal matching and imitation for the development of representation and memory in infancy. *Annals of the New York Academy of Sciences*, *608*, 1–31.

Ohala, J. J. (1994). The frequency code underlies the sound-symbolic use of voice pitch. In L. Hinton, J. Nichols, & J. J. Ohala (Eds.), *Sound symbolism* (pp. 325–347). Cambridge, England: Cambridge University Press.

Ohala, J. J. (1997). Sound symbolism. In *Proceedings of the 4th Seoul International Conference on Linguistics [SICOL]* (pp. 98–103). Seoul, South Korea: Linguistic Society of Korea.

Sapir, E. (1929). A study in phonetic symbolism. *Journal of Experimental Psychology*, *12*, 225–239.

Shi, R., Werker, J. F., & Morgan, J. L. (1999). Newborn infants' sensitivity to perceptual cues to lexical and grammatical words. *Cognition*, *72*, B11–B21.

Shultz, T. R. (2007). The Bayesian revolution approaches psychological development. *Developmental Science*, *10*, 357–364.