Auditory-visual crossmodal numerosity matching by infants and young children

Tessei Kobayashi, Kazuo Hiraki, and Toshikazu Hasegawa
The University of Tokyo*

This study investigated whether infants and young children are able to perform crossmodal matching of small numerosities, i.e., the recognition of numerical equivalence between sets presented with different sensory modalities. In an infant experiment that used a violation-of-expectation paradigm, 6-month-olds looked significantly longer at the numerically non-equivalent events (e.g., 2-tone/3-object events) than at the numerically equivalent events (e.g., 3-tone/3-object events). In a child experiment that used a matching-to-sample task, 3- to 4-year-old children were able to select correctly the stimulus card that matched the number of tones they heard (0–4 tones). These findings suggest that infants and children are capable of relating small numerosities of sets presented with different sensory modalities.

Key words: cognitive development, numerical cognition, crossmodal matching

There is still debate over whether early numerical ability is modality independent (e.g., Kobayashi et al., 2004). In order to address this issue, the present study investigated whether infants and young children are able to perform auditory-visual crossmodal matching of small numerosities (up to 3 or 4), i.e., the recognition of numerical equivalence between sets presented with different sensory modalities.

Experiment 1

The participants were 32 full-term infants with a mean age of 6 months, 4 days (range: 5 months, 10 days to 6 months, 30 days). Each infant sat on the lap of a parent at approximately 150 cm from a screen (66.7 × 50 cm) on which a color, computer-animated movie was presented by using a multimedia projector. A violation-of-expectation paradigm was used. The infants were first shown 4 familiarization trials in which 2 or 3 objects (animated characters like Mickey Mouse) impacted a surface with a tone. They were alternatively presented with the 2-object and the 3-object sequences with each of the sequences presented twice. While the movement of the objects was visible in the 1st and 2nd trials, an opaque screen blocked the lower half of their movement in the 3rd and 4th trials. In the test trials the infants were presented with 2 and 3 tones while the movement of each object remained hidden behind an opaque screen. The infants in one task were presented with auditory tones at a constant rate (1 tone per 4 sec). The infants in the other task were presented with auditory tones at a constant duration (a total duration of 8 sec). Within each task the infants were assigned to one of two conditions. The infants in one condition (the Two-object Condition) watched the 2-object display in every trial after being alternatively presented with the 2-tone (expected event) or the 3-tone sequences (unexpected event). Those in the other condition (the Three-object Condition) watched the 3-object display in every trial after being alternatively presented with the 2-tone (unexpected event) or the 3-tone sequences (expected event). All of the infants were shown 4 test trials with each of the events presented twice. A recording of an infant's viewing of each trial began when the objects appeared, and finished when the infant either looked away from the display for 2 consecutive seconds or had viewed the screen for 30 sec.

A repeated-measures ANOVA with Task (Constant Rate vs. Constant Duration) and Condition (Two-object vs. Three-object) as the between-subjects factors, and Event (Expected and Unexpected) and Test-
trial Pair (First and Second) as the within-subjects factors yielded significant main effects of Event, $F(1, 28) = 53.505, p = 0.0001$, and Test-trial Pair, $F(1, 28) = 16.733, p = 0.0003$. There were no other significant main effects and no significant interactions. These results showed that the infants in both conditions looked significantly longer at the unexpected events (Constant Rate; Two-object = 20.2 sec, Three-object = 16.0 sec; Constant Duration; Two-object = 16.0 sec, Three-object = 14.7 sec) than at the expected events (Constant Rate; Two-object = 13.5 sec, Three-object = 10.6 sec; Constant Duration; Two-object = 10.6 sec, Three-object = 8.7 sec) irrespective of either task (see Figure 1). The results suggested that the infants seemed more likely to form an expectation about the numerosity of visual objects behind the screen based on the sequences of the auditory tones they heard.

**Experiment 2**

Three groups of 48 children participated in the study. The ages of the groups were as follow: 3-year-olds (mean age, 3 years 1 month; range, 2 years 10 months to 3 years 4 months); 3-1/2-year-olds (mean age, 3 years 10 months; range, 3 years 7 months to 4 years); and 4-year-olds (mean age, 4 years 6 months; range, 4 years 3 months to 4 years 10 months). Each group had the same number of boys and girls. A matching-to-sample (MTS) procedure was used. Each child was seated in a chair, approximately 60 cm from a 14.1-inch laptop computer on a table. A color computer-animated movie was presented on the monitor. The child was first presented with the sequences of auditory tones (0, 1, 2, 3, and 4) as a sample stimulus while the movement of each object remained hidden behind an opaque screen. Then, the child was presented with the 5 stimulus cards that showed different numbers of visual objects (0 to 4). The cards consisted of horizontally arrayed sets of objects. Next, the child was required to choose one stimulus card that matched the sample in terms of numerosity while the objects still remained hidden. The child was shown 10 test trials with each of 5 types of auditory sequences presented twice. The order of the sequences was counterbalanced across the partici-

![Figure 1. Mean looking times of expected and unexpected events in each condition or task in Experiment 1.](image)

pants. A correct choice of the matching card was scored with 1 point, resulting in a total score of 10 for each child.

The results showed that the children in all of the groups performed the MTS task significantly above chance. Their results were as follow: 3-year-olds, $M = 5.69, t (15) = 5.563, p < 0.0001$; 3-1/2-year-olds, $M = 6.50, t (15) = 6.903, p < 0.0001$; 4-year-olds, $M = 7.88, t (15) = 10.192, p < 0.0001$. The findings suggest that 3- to 4-year-old children are able to perform cross-modal matching of small numerosities correctly.

**Discussion**

Taken together, these findings suggest that infants and young children seem more likely to form an expectation about the numerosity of visual objects behind a screen based on the sequence of auditory tones they heard. That is, they seem to relate small numerosities of sets presented with different sensory modalities. Therefore, they may possess a modality independent representation for numerosity.

**References**