Reprint Series
18 March 1983, Volume 219, pp. 1347–1349

SCIENCE

Infant Intermodal Speech Perception Is a Left-Hemisphere Function

Kristine MacKain, Michael Studdert-Kennedy, Susan Spieker, and Daniel Stern

Infant Intermodal Speech Perception Is a Left-Hemisphere Function

Abstract. *Prelinguistic infants recognized structural correspondences in acoustic and optic properties of synchronized, naturally spoken disyllables, but did so only when they were looking to their right sides. This result suggests that intermodal speech perception is facilitated by rightward orientation of attention and subserved by the left hemisphere.*

Research on infants' capacities for intermodal perception has demonstrated that infants are sensitive to correspondences in the acoustic and optic properties that specify an event (*1*, *2*). Infants may prefer a natural pattern of structural correspondence between the optic and acoustic dimensions of an event by which, in speech for example, an opening mouth is correlated with a rise in amplitude and with an upward shift in overall spectral structure and a closing mouth with the reverse. Alternatively, infants may simply prefer a temporal pattern of correspondence by which gross points of change in acoustic and optic structure are synchronized (*1*). If infants prefer mere synchrony, they should be satisfied with any arbitrary pattern of acoustic-optic correspondence; thus, in speech they might have no preference for syllable amplitude peaks synchronized with an open mouth over those synchronized with a closed mouth. But if infants prefer natural patterns of structural correspondence, they should look longer at the synchronized video monitor display of a woman producing articulatory patterns that specify the speech they are hearing than at an alternative, synchronized video display of the same woman displaying a different articulatory pattern. We therefore investigated the capacity of infants to recognize acoustic-optic correspondences in speech structure when the synchrony between an acoustic and two competing optic displays was maintained.

Our preliminary analyses suggested that when acoustic and optic speech displays specified the same disyllable, intermodal recognition was enhanced if infants were watching the right, rather than the left, video display. When adults look to the right (or left) as they complete a task, their performance is facilitated if the task demands are better subserved by the hemisphere contralateral to gaze direction (*3*). Such results have been interpreted as evidence that attention, behaviorally manifested by gaze, may selectively activate the hemisphere contralateral to direction of gaze. We therefore expected that only rightward looking would significantly enhance recognition of acoustic-optic correspondences in speech structure.

Eighteen infants, eight males and ten females, 5 to 6 months of age ($\bar{X}$ = 5 months, 25 days) participated in the experiment. We used three pairs of naturally produced consonant-vowel-consonant-vowel (CVCV) disyllables, spoken with equal stress on both syllables: /mama lulu/, /bebi zuzi/, and /vava zuzu/. We enhanced the opportunity to detect acoustic-optic correspondences by making the articulatory dynamics of the contrasting video displays highly discriminable. To prepare the experimental materials, an adult female silently articulated each CVCV in synchrony with either the corresponding or the contrasting spoken disyllables of another adult female. The voice and the articulating face were recorded simultaneously to appear on one side of a 28 cm by 22 cm video monitor screen. The video recording procedure was then repeated so that the articulating face appeared on the other half of the split video screen, silently articulating the second CVCV in the pair in synchrony with the audio playback of the original disyllable. Deviations in acoustic-optic synchrony were below the adult threshold for detecting asynchronies (*4*). The resulting recording of the acoustic signal synchronized with two competing articulatory displays was output to two video monitors.

The infant sat 46 cm from the video monitors on its mother's lap at the open end of a wooden box. The infant viewed a different articulatory display on the split screen of each monitor, one appearing through the right back window of the box and the other through the left. The

Table 1. First fixation times in seconds, averaged across six disyllables, to the left and right video display when the display matched or mismatched the audio CVCV. Mean fixation times are summed across 18 infants.

Direction of gaze	Video display	
	Matches audio CVCV	Mismatches audio CVCV
Left	66.0	59.3
Right	81.2	67.0

Table 2. Proportion of first fixation time, averaged over 18 infants, spent looking at right matches or mismatches, left matches or mismatches, and right or left matches.

Proportion of time spent looking at	Disyllable						Overall
	bebi	zuzi	mama	lulu	vava	zuzu	
Matches versus mismatches							
Right	.59	.52	.62	.53	.52	.61	.57
Left	.54	.50	.54	.49	.49	.52	.51
Right versus left matches	.57	.57	.61	.52	.58	.59	.57

speech corresponding to one of the two video displays was played at equal loudness from speakers of both monitors. A camera placed centrally between the monitors taped the infant's visual responses. The mother looked over the roof of the box and could not see the video displays.

Infants were presented with each of the three CVCV pairs on four trials for a total of 12 trials. Each member of a CVCV pair occurred twice as an audio signal, with its matching video display occurring once on the left video monitor and once on the right. The trials were randomized under the constraint that no two trials with the same video output immediately follow one another. We used nine different randomizations and assigned two infants to each. Each trial lasted 20 seconds and consisted of 11 auditory-visual CVCV repetitions. Disyllable durations were about 1100 msec, separated by interstimulus intervals of about 800 msec. Successive trials began without interruption between trials. The experimental session lasted 4 minutes.

From video recordings of the child's face, independent observers recorded for each trial the duration in seconds of the first fixations to the right and to the left. We preferred first fixation over total fixation time because it is less vulnerable to contamination by factors such as attentional lapse. Interjudge reliability, based on a Pearson product-moment correlation coefficient for 41 randomly selected trials, was $r = .96$ for left-looking time and $r = .98$ for right-looking time.

The direction of the infants' first looks after trial onset was to the right side on 58 percent of the total trials ($N = 216$). The longest first fixation times were to matches, particularly on the right side (Table 1).

Because first fixation times varied across infants, we obtained proportions of first fixation time spent looking at acoustic-optic matches occurring on the right and the left side by each infant for each disyllable. We thus normalized for variability over subjects and disyllables and, at the same time, for any general preference for one side over the other. Proportions were computed by dividing the first fixation time spent looking at a match (right, left, or both sides) by the total first fixation time for that comparison, summed across two trials (Table 2).

The overall proportion of total (right and left) first fixation time spent looking at matches ($\bar{X} = 0.54$) rather than mismatches was significant ($z = 2.64$, $P < .004$, $N = 16$, two ties; this and subsequent tests are one-tailed Wilcoxon matched-pairs signed-ranks tests). Table 2 summarizes the remaining results.

On the right side, the proportion of first fixation time spent looking at matches was significantly greater than for mismatches overall ($z = 2.66$, $P < .004$, $N = 18$) and for three of the six disyllables: *mama*, *bebi*, and *zuzu* (with respective values of $z = 2.46$, $P < .007$, $N = 17$, one tie; $z = 1.94$, $P < .03$, $N = 17$, one tie; $z = 2.27$, $P < .01$, $N = 18$). Proportions were greater than .50 for all six disyllables. On the left side, the proportion of first fixation time spent looking at matches was not significantly greater than for mismatches overall or on any of the six disyllables. Proportions were greater than .50 for only three of the disyllables.

On the right side, the number of infants who spent more than half of their first fixation time looking at matches versus mismatches was significant, on a binomial test, for two disyllables (*mama*, 13/18, $P < .05$; *zuzu*, 14/18, $P < .02$), but no corresponding tests for left-side looking were significant.

In a right-left comparison, the proportion of first fixation time spent looking at acoustic-optic matches was significantly greater on the right side than on the left side overall ($z = 2.02$, $P < .02$) and for three out of the six disyllables: *mama*, *bebi*, and *zuzu* (respectively, $z = 1.87$, $P < .03$, $N = 17$, one tie; $z = 1.68$, $P < .05$; $N = 18$; $z = 1.96$, $P < .03$, $N = 18$). Proportions on the right side were greater than those on the left for all six disyllables (Table 2).

One potential source of bias—a preference for an optic articulatory pattern irrespective of the acoustic pattern that accompanied it—might have influenced these results. To check for this, Spearman rank-order correlation coefficients were computed for preferences for a video display when the audio signal matched the video display and when it did not. We computed correlations for right and left sides combined as well as for each side separately. A significant positive correlation would have indicated that infants preferred to look at a particular articulatory pattern irrespective of the CVCV to which they were listening; none of the correlations was significant.

Because infants looked significantly longer at synchronized video displays of a woman articulating a disyllable synchronized and matched with what they were hearing than at an alternative display synchronized but not matched with what they were hearing, their preference was for acoustic-optic correspondences in structure, not for mere synchrony. Moreover, they displayed this preference only when attending to the right side.

These findings demonstrate (i) sensitivity of infants to natural structural correspondences, rather than merely temporal ones, between the acoustic and optic properties of articulation and (ii) mutual facilitation of two left-hemisphere functions: rightward orientation of attention (*3*) and intermodal speech perception. Taken with the well-known dominance of the left hemisphere in the motor control of speech for adults (*5*) and in speech perception for both adults (*6*) and infants (*7*), these results suggest that the normal infant's capacity to begin reproducing native language speech sounds in prelinguistic babbling (*8*), may rest on a predisposition of the left hemisphere to recognize the sensorimotor connections between the auditory structure of speech and its articulatory source.

KRISTINE MACKAIN
Department of Psychiatry, Cornell University Medical College, New York 10021
MICHAEL STUDDERT-KENNEDY
Department of Communication, Arts, and Sciences, Queens College, and *Graduate Center, City University of New York,* and *Haskins Laboratories, New Haven, Connecticut 06510*
SUSAN SPIEKER*
DANIEL STERN
Department of Psychiatry, Cornell University Medical College

References and Notes

1. E. Spelke, *Dev. Psychol.* **15**, 626 (1979).
2. ——— and A. Cortelyou, in *Infant Social Cognition: Empirical and Theoretical Considerations*, M. E. Lamb and L. R. Sherrod, Eds. (Erlbaum, Hillsdale, N.J., 1981), pp. 61–84; B. Dodd, *Cognit. Psychol.* **11**, 478 (1979).
3. M. Kinsbourne, *Acta Psychol.* **33**, 193 (1970); in *Attention and Performance V*, R. M. A. Rabbitt

and S. Dornic, Eds. (Academic Press, London, 1974), pp. 81–97; H. Lempert and M. Kinsbourne, *Neuropsychologia* **20**, 211 (1982).
4. Temporal discrepancies in audio-video speech events must reach 131 msec before they can be detected by adults [N. Dixon and L. Spitz, *Perception* **9**, 719 (1980)]. In our study, temporal discrepancies between corresponding events on any two video displays did not exceed 48 msec. Furthermore, there were no significant differences in seven adults' perceptual judgments of temporal discrepancies between acoustic-optic matches versus mismatches for any of the six disyllables. We assumed that infants' sensitivity would not be superior to adult's on this task. The procedures have been detailed in a paper presented at the 2nd International Conference for the Study of Child Language, Vancouver, B.C., 9 to 14 August 1981.
5. B. Milner, in *The Neurosciences: Third Study Program*, F. O. Schmitt and F. G. Worden, Eds. (MIT Press, Cambridge, Mass. 1974), pp. 75–89.
6. M. Studdert-Kennedy and D. Shankweiler, *J. Acoust. Soc. Am.* **48**, 579 (1970).
7. D. L. Molfese, R. B. Freeman, D. S. Palermo, *Brain Lang.* **2**, 356 (1975); C. T. Best *et al.*, *Percept. Psychophys.* **31**, 75 (1981).
8. B. de Boysson-Bardies, L. Sagart, N. Bacri, *J. Child Lang.* **8**, 511 (1981).
9. We thank A. Liberman and B. Repp for critical comments and J. Monroe and B. Repp for statistical advice. Supported in part by NICHD postdoctoral fellowship HD-05407 to K.M., by a grant from the Jane Hilder Harris Foundation to Cornell University Medical College, and by NICHD grant HD-01944 to the Haskins Laboratories.
* Present address: Child Development and Mental Retardation Center, University of Washington, Seattle 98195.

23 June 1982; revised 30 September 1982