Child Development, September/October 2000, Volume 71, Number 5, Pages 1241–1257

Infants' Perception of the Audible, Visible, and Bimodal Attributes of Multimodal Syllables

David J. Lewkowicz

Three experiments investigated 4-, 6-, and 8-month-old infants' perception of the audible, visible, and combined attributes of bimodally specified syllables. Ninety-six infants in each experiment were habituated to a person mouthing and uttering a syllable and then tested for detection of changes of either the audible, visible, or combined attributes of the syllable. When the attributes of the syllable were produced in an adult-directed manner, all three age groups discriminated the audible and bimodal attribute changes but only the 8-month-olds discriminated the visible one. When the difference between the familiar and novel attributes of the syllable was enhanced by testing with a novel syllable produced in an infant-directed manner, all three age groups detected all three types of changes. Finally, to test the possible role of audiovisual synchrony in responsiveness, infants were tested with an asynchronous syllable spoken either by the same person or by a novel person following habituation to a synchronous syllable. Results suggested that at four months infants attended primarily to the featural information, at six months primarily to the asynchrony, and at eight months to both features independently. These results help identify some of the important dimensions of multimodal speech during early development.

INTRODUCTION

Interactions with caregivers, siblings, and others constitute a frequent and ubiquitous part of an infant's everyday life. During such interactions infants usually not only can hear another person speaking but can see that person at the same time. In other words, during most of their interactions with social partners infants not only listen to them but see them as well. This raises an interesting question from a perceptual development point of view: How do the audible and visible components of a talking face contribute separately and together to speech perception in early development? Historically, research investigating perceptual development in infancy has tended to focus on single sensory modalities and, as a result, asking questions about the relative contribution of multiple sensory inputs has not been of primary interest to researchers. This is surprising given that the theoretical importance of intersensory interactions for perception, perceptual development, and the acquisition of higher level skills has been repeatedly noted (Birch & Lefford, 1963, 1967; Edelman, 1992; Gibson, 1966, 1969, 1984; Maier & Schneirla, 1964; Piaget, 1954; Thelen & Smith, 1994; Werner, 1973).

In fact, multimodally specified events can produce perceptual effects that are qualitatively different from unimodally specified ones. One of the most dramatic examples illustrating this fact comes from research on adults' perception of speech. In a classic study, McGurk and MacDonald (1976) showed that the auditory perception of a syllable can be altered by a conflicting visual syllable. For example, when McGurk and MacDonald presented an auditory /ba/ and a visible /ga/ and asked subjects to identify what they heard, they reported hearing the syllable (/da/). In other words, responsiveness to an audible syllable is completely different depending on whether it is presented unimodally or in the context of input in a different sensory modality. The ventriloquism effect is another example of the importance of multimodal information for perception. In this case, when subjects are presented with a lateralized sound and a concurrent but spatially disparate visual stimulus, they report hearing the sound as emanating from a location closer to the visual stimulus (Bertelson & Radeau, 1976, 1981; Jack & Thurlow, 1973). The McGurk and ventriloquism effects show that the sensory modalities usually work in concert. This tendency to integrate information across modalities is so powerful that when the information available to the different sensory modalities conflicts, our perceptual system either produces completely novel percepts or ones that are a compromise between the two sources of information. These two examples make a convincing case that studying responsiveness to multimodal objects and events is important. In addition, however, they lead to one nonobvious conclusion: If one wishes to study the perception of multimodally specified events, one must study responsiveness to the information in one sensory modality *in the context* of the information in the other modality.

Most of the research on infant intersensory percep-

0009-3920/2000/7105-0013

tion has focused on the detection of intersensory equivalence (or intermodal invariants) partly because of the influence of the direct perception view (Gibson, 1969) and partly because the detection of intersensory equivalence is the most obvious of the various forms of intersensory interactions (Marks, 1978). This work has shown that intersensory interactions do occur early in human development (Lewkowicz & Lickliter, 1994). For example, it has been shown that neonates can equate auditory and visual inputs on the basis of intensity (Lewkowicz & Turkewitz, 1980) and that older infants can perceive audiovisual (AV) relations based on synchrony and duration (Bahrick, 1994, 1996; Lewkowicz, 1986, 1992a, 1992b, 1996b; Walker, 1982) and can perceive the AV aspects of affect (Walker-Andrews, 1986) and speech (Kuhl & Meltzoff, 1982). Indeed, there is even evidence that infants exhibit responsiveness consistent with the kinds of effects predicted by the McGurk effect (Burnham, 1998; Rosenblum, Schmuckler, & Johnson, 1997).

Despite the fact that the detection of intersensory equivalence is an important aspect of intersensory interaction, some have argued that "The various senses are not simply equivalent ways of perceiving important classes of events; they differ in both the precision and rapidity of their action." (Welch & Warren, 1986, p. 2). Therefore, the question of the relative role of multimodal inputs on infant perceptual response should not be ignored. In fact, this question is particularly important when considering intersensory perception during early development because the detection of intersensory equivalence has to occur against constantly changing sensory, perceptual, and cognitive capacities (Kellman & Arterberry, 1998). For example, the different sensory modalities have their functional onset at different times and follow different developmental schedules (Gottlieb, 1971). As a result, one sensory modality might be ahead of the other in its overall functional superiority. Alternatively, the relative functional superiority of one modality over the other might depend on the specific nature of the stimulation presented, which may, in turn, be the result of relative developmental sensory experience (Lewkowicz, 1999).

A good deal is now known about infants' responsiveness to the audible and visible components of talking faces (Walker-Andrews, 1997). At the same time, however, questions about the absolute degree to which the vocal and facial features of a talking face contribute to perception are virtually impossible to answer because equating stimulation in different modalities a priori is very difficult. One way around this problem is to use a convergent-operations approach in which the same experimental method is used across a series of studies to investigate responsiveness to the components of multimodal compound stimuli. One such method is the multimodal component variation method used by Lewkowicz (1988a, 1988b, 1996a, 1998) to study infant responsiveness to the audible and visible components of multimodal compounds. To study responsiveness to face/voice compounds, the method involves first habituating infants to a compound stimulus consisting of the videotaped image of a person who can be seen and heard reciting a prepared script. Following habituation, each infant is given a series of test trials during which some aspect of either the audible, visible, or combined components of the talking face is changed. If the infant exhibits response recovery when a given component is changed, then this is interpreted to mean that she perceived the change. One of the key features of this method is that it incorporates the requirement that one should study responsiveness to input in one modality in the context of input in the other.

To study infants' responsiveness to talking faces, Lewkowicz (1996a) habituated 4-, 6-, and 8-month-old infants to a videotaped segment of a female actor reciting a continuous utterance spoken in an adult-directed (AD) manner (the way one adult would normally speak when speaking to another adult). Following the habituation phase infants were given a set of test trials. During the auditory (A) test trial a novel utterance spoken by a novel person could be heard but the familiar face mouthing the familiar script could still be seen; during the visual (V) test trial a novel face could be seen mouthing a novel script but the familiar person uttering the familiar script could be heard; during the auditory/visual (A/V) test trial a novel face could be seen and heard uttering a new script. In addition to these three test trials, a posttest trial was given during which a completely novel stimulus (a segment of a "Winnie-the-Pooh" cartoon) was presented to determine whether the failure to respond in the preceding test trials resulted from failure to discriminate the specific changes or to fatigue.

Results indicated that for AD speech only the 6- and the 8-month-old infants detected the changes and only the bimodal and visible ones. These findings were interpreted to mean that infants between 4 and 8 months of age have difficulty discriminating the audible features of continuous bimodal AD speech but that beginning at 6 months they can discriminate its visible attributes. Given that infants in this age range already possess rather impressive speech perception abilities for unimodal speech (Jusczyk, Houston, & Goodman, 1998) it was rather surprising that the infants did not respond to the audible change. As was noted earlier, however, responsiveness to unimodal stimulation can differ from responsiveness to the same input in a multimodal context. Moreover, detection of differences in continuous utterances may differ from infants' detection of isolated and simple phonemic contrasts of the type usually studied.

To explore possible reasons for the failure of infants to discriminate the audible features of the talking face, Lewkowicz (1996a) conducted two additional studies in which either a gender difference or a manner-of-speech difference was introduced. Both of these types of differences are known to be discriminable to infants in this age range (Fernald, 1985; Miller, 1983). Introduction of a gender difference did not aid in the discrimination of the audible change at any age but did make it possible for the 4-month-old infants to detect the bimodal and visible changes. Introduction of a manner-of-speech difference on top of a gender difference (i.e., by habituating infants to AD speech produced by a male and testing with aspects of infant-directed (ID) speech produced by a female) did, however, make it possible for the 6- and 8-month-old infants to detect the audible as well as the bimodal and visible changes. The 4-month-old infants, however, still failed to respond to the audible changes even though they responded to the visible and combined ones. These findings show that 6- and 8-month-old infants are sensitive to manner of speech in a bimodal context when those changes involve a change from AD to ID speech along with a gender change. They suggest that the exaggerated prosodic variations associated with ID speech can contribute to infants' perception of the audible attributes of bimodal continuous speech.

Given that prosodic variations of the type found in ID speech appeared to aid in the discrimination of the audible aspects of bimodal speech, Lewkowicz (1998) conducted two further studies in 3-, 4-, 6-, and 8-month-old infants to explore the development of sensitivity to other types of prosodic variations. In one study, responsiveness to the difference between a person speaking and another person singing was investigated. In the second study, responsiveness to speech and singing coupled with gender differences was tested. When gender differences were absent, 4-, 6-, and 8-month-old infants discriminated all three kinds of changes whereas the 3-month-old infants discriminated only the bimodal and visible changes. The fact that the 4-month-old infants responded to the audible change was especially interesting because it showed that when more exaggerated prosodic variations associated with singing are provided, infants can discriminate the difference between the audible attributes of continuous speech and singing earlier in development. Results from the second study showed that when gender was added to the speech-singing difference, all four age groups discriminated all three types of changes.

Although these findings offer some insight into what stimulus features might be important for infants' perception of the audible attributes of continuous speech in the multimodal context, they also raise an interesting question. As noted earlier, one of the findings from these studies was that infants did not discriminate the audible differences in AD speech even at 8 months of age. What makes this finding intriguing is that this was the case even though the audible changes involved differences in the identity of the voice and the nature of the script. This meant, in turn, that there were concurrent changes in content, the temporal distribution of the content, and the spectral characteristics of the voice. In addition, the utterances consisted of a hierarchy of distinct units such as uni- and multisyllabic words, phrases, and clauses. In other words, infants had a set of potentially useful discriminative cues available to them but did not take advantage of them. One possible reason for the infants' failure to take advantage of these discriminative cues might be the overly complex nature of the auditory information or its relationship to the concurrent visual information. To test this possibility, in the current set of experiments we investigated infants' detection of A, V, and A/V differences in simpler (unisyllabic) utterances. In addition, we examined the influence of type of speech (AD, ID) and featural information on the detection of such differences and the role that AV temporal relations play in responsiveness to unisyllabic utterances. By examining infants' responsiveness to face/voice compounds representing the other end of the complexity continuum, we hoped to provide a fuller picture of the way infants perceive them and to get greater insight into the role of complexity.

EXPERIMENT 1

We used the multimodal component variation method to test infants' responsiveness to the A, V, and A/V attributes of single syllables by habituating them to a videotaped segment of a female who could be seen and heard repeatedly reciting a single syllable in the AD manner. Once infants reached a habituation criterion, they were given three types of test trials during which the same person also spoke in the AD manner except that now the infant either heard a novel syllable while seeing the familiar one, saw a novel syllable while hearing the familiar one, or heard and saw a novel syllable. To maximize the contrast between the familiar and novel syllables, the two syllables that were chosen for this experiment were clearly distinct in terms of both their audible and visible properties; they differed in terms of their voicing, place of articulation, and manner of articulation. On the basis of the fact that prior studies (Lewkowicz, 1996a, 1998) have shown that there are important developmental differences in the way infants respond to the audible and visible attributes of talking faces, three different age groups were tested.

Method

Participants. The sample for this experiment consisted of a total of 96 participants. It included separate groups of 32 infants each, aged 4, 6, and 8 months of age with an equal number of boys and girls at each age. The mean age of the 4-month group was 18.9 weeks (SD = 1.1 weeks). One additional 4-month-old infant was tested but this infant's data were eliminated because of sleepiness. The mean age of the 6-month group was 27.6 weeks (SD = .7 weeks). Four additional 6-month-old infants were tested but their data were eliminated because three of them were fussy and one was distracted by the parent. The mean age of the 8-month group was 36.7 weeks (SD = .8 weeks). One additional 8-month-old infant was tested but this infant's data were eliminated because of fussiness. Infants for this experiment as well as for the subsequent two experiments were recruited from the birth records of a local hospital and were term at the time of birth, had birthweights greater than 2,500 g, had 1- and 5-min Apgar scores greater than 7, and were healthy at the time of testing. The infants in all three experiments were primarily of middle-class background and predominantly European American.

Apparatus and stimuli. Each infant sat either in an infant seat, a high chair, or a parent's lap in front of a 13-inch color video monitor (Panasonic CT-1331Y) located at a distance of 50 cm from the infant. Infants who were fussy when first placed in the infant seat or high chair were placed on the parent's lap. Only two 6-month-old and four 8-month-old infants were tested on the parent's lap (the parent, who was unaware of the specific purpose of the experiment, was asked to sit as still as possible, and refrain from interacting with the baby). A color camcorder (Panasonic AG-190), located on top of the stimulus display monitor, was used to videotape the infant's behavior and to enable the experimenter to view the baby during the experiment. A curtain, extending from the monitor past the point where the infant was seated, was placed on each side to occlude the infant's peripheral view.

Six different stimulus events (video/audio recordings) were first videotaped and then transferred to a 12-inch laser disk. Four of these stimulus events showed one of two female actors reciting one of two syllables; in two of these one of the actors could be seen and heard repeating the syllable /ba/ or /sha/ and in the other two the other actor could be seen and heard repeating the same two syllables. The fifth event, the "attention-getter," was a silent, green disk that alternately expanded and shrank. It measured 8 cm in diameter at its largest size and 2 cm at its smallest size. The sixth and final stimulus event was a segment of a "Winnie-the-Pooh" cartoon that served as a pre- and a posttest trial.

Each of the four recordings portrayed the actor such that the top and bottom of her head corresponded to the top and bottom edges of the stimulus display monitor, respectively. During filming the actor was asked to repeat the syllable at a relatively steady pace and to speak as if she were speaking to another adult (i.e., in the AD manner). This meant that while speaking, she held her head still, recited the syllable with a minimum amount of intonation, and did not smile. The interval between each instance of the syllable ranged between 2 and 3 s. The average sound intensity was 71 dB (A scale, re. .0002 dynes/cm^2) while one actor recited the two syllables and 73 dB while the other actor recited the syllables, with a range of approximately 4 dB in each case (as measured by a Bruel and Kjaer 2203 sound-level meter with the microphone placed at the same place as the infant's head). After the four syllable events and the "Winnie-the-Pooh" segment were recorded, they were reviewed and 34-s segments were chosen and recorded onto a second videotape. The contents of this videotape then were transferred to the laser disk. The same was done with the attention-getter display except that its duration was several minutes long so as to permit infants sufficient time to reorient toward the screen before the start of each trial. The average intensity of the pre- and posttest segments was 71 dB with an 8-dB range. A Macintosh IIci computer connected to a Sony LDP-1550 laser-disk player was used to control all stimulus presentations and to record responsiveness.

Procedure and design. The experimenter watched the infant on a television monitor that was connected to the camcorder and controlled stimulus presentation by pressing a key on the computer keyboard to initiate a trial whenever the infant looked at the stimulus display monitor. Stimulus access on the laser disk was very rapid (between 1 and 2 s). As long as the infant looked at the stimulus display monitor, the experimenter continued to hold down the keyboard button and thus stimulus presentation continued. As soon as the infant looked away, the experimenter ceased pressing the button. The computer program controlling the experiment automatically recorded the duration of each button press and thus the duration of looking for that trial. Although the recording of visual attention as a measure of responsiveness to auditory stimulation may not seem obvious, prior studies (Horowitz, 1974) have shown that visual attention to changing auditory stimulation can provide a reliable and useful measure of auditory processing. This is the case for studies of infants' responsiveness to the auditory component of an auditory/visual compound stimulus (Lewkowicz, 1988a, 1988b, 1992b, 1996a), as well as for studies of infants' responsiveness to an auditory stimulus alone (Jusczyk & Kemler Nelson, 1996).

An infant-controlled habituation/test procedure was used. Each trial began with the presentation of the attention-getter in the center of the stimulus display. As soon as the infant looked at it, the attention-getter was turned off and one of the stimulus segments was presented. The segment continued for as long as the infant looked at the monitor, up to a maximum of 34 s. The upper limit for trial length was determined by the amount of stimulus material that could be contained on a single laser-disk and still accommodate all possible types of trials. Whenever the infant either looked away from the monitor for more than 1 s or 34 s was reached, the stimulus was turned off (this signaled the end of the trial) and the attention-getter reappeared. As soon as the infant looked back at the monitor, the attention-getter disappeared and a stimulus segment was presented again. Throughout the session, the experimenter listened to music presented through a pair of headphones and was not able to see the stimuli being presented. As a result, the experimenter was blind with respect to the audible and visible characteristics of the stimulus being presented.

For each infant, an experimental session consisted of a pretest trial, the habituation trials, four test trials, and a single posttest trial in that order. The novelty of the material presented during the pre- and posttest trials permitted an assessment of fatigue effects and made it possible to determine whether fatigue might account for any failures to exhibit response recovery during the test trials. At each age, equal numbers of infants ($n = 8$) were habituated to one of the four combinations of actor × syllable stimulus events.

Habituation was deemed complete when the total amount of looking during the last three trials decreased by 50% relative to the total amount of looking during the first three trials. This criterion was calculated by way of a moving window such that the second set of three trials began with the second trial and ended with the fourth trial. If the habituation criterion was not met at this point, the window slid down one trial and included the third through the fifth trials. This sliding window criterion continued to be applied until the 50% criterion was met. The four test trials consisted of a familiar (F) test trial during which an infant saw and heard the same person that she heard and saw during the habituation phase, an A test trial during which the infant heard the familiar person speaking a novel syllable but saw her mouthing the familiar syllable, a V test trial during which the infant heard the familiar person speaking the familiar syllable but saw her mouthing a novel syllable, and an A/V test trial during which the infant heard and saw the familiar person speaking a novel syllable. (It should be noted that during the A and V test trials the audible and visible information was not in synchrony because the lips were seen producing something different than the voice was heard producing. This is unavoidable with this type of design. Experiment 3 was designed specifically to further address this issue and ask whether synchrony played a role in responsiveness.) The four test trials were presented according to one of four orders across infants in each of the four person × syllable habituation subgroups at each age. These four orders were chosen so that each of the four test trials was presented equally often in each ordinal position in each subgroup at each age.

Results and Discussion

The results from the pretest, test, and posttest trials are depicted in Figure 1. The first question that was addressed by two related analyses was whether infants became fatigued during the session. The first of these analyses compared the results from the pretest and posttest trials by way of a 2 (trials) × 3 (age) × 2 (gender) ANOVA. This analysis showed that infants looked for an average of 19.4 s during the pretest trial and an average of 18.6 s during the posttest trial and that this was not a significant difference (neither gender nor age were significant factors). The second of these analyses compared the amount of looking in the posttest trial versus the amount of looking in the F test trial ($M = 8.4$ s) and indicated that infants recognized the novelty of the posttest stimulus, $F(1, 90) = 104.12$, $p < .001$ (neither gender nor age contributed to this effect). These findings indicate that any failure to exhibit significant response recovery during the test trials could not have been due to fatigue effects because all of the test trials preceded the posttest trial.

The purpose of the principal analysis was to determine how infants responded to the various component changes. To determine whether age, gender, or type of habituation segment affected outcome, an initial analysis examined responsiveness in the F, A, V, and A/V test trials by way of a mixed ANOVA, with

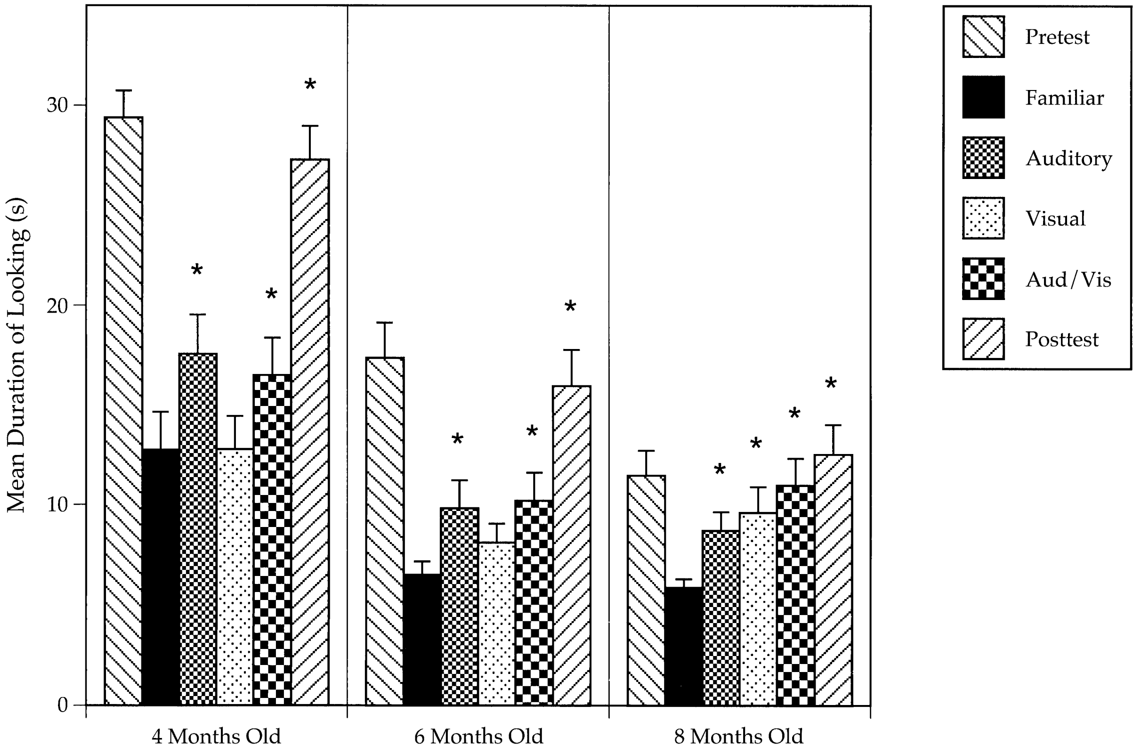

Figure 1 **The results from Experiment 1 showing the mean duration of looking plotted separately for each age group during the pretest, test, and posttest trials. Error bars signify standard error of the mean and asterisks indicate significant response recovery.**

3 (age), 2 (gender), and 4 (person × type of syllable combination) as between-subjects factors and 4 (trials) as a within-subjects factor. Results of this analysis yielded a significant effect of trials, $F(3, 216) = 9.5$, $p < .001$, and a significant age effect, $F(2, 72) = 11.98$, $p < .001$. Although this analysis did not yield a significant Age × Trials effect, when the data from the posttest trial were included as part of the trials factor, the Age × Trials effect was significant, $F(8, 288) = 3.52$, $p < .001$.

Given the significant Age × Trials interaction, the test trial data were analyzed separately for each age group. This was done by conducting separate sets of planned contrast analyses at each age to determine whether infants discriminated the various component changes. The planned contrasts compared the duration of looking in the F test trial with the amount of looking in the A, V, and A/V test trials, respectively. As can be seen in Table 1, the 4-month-old and the 6-month-old infants exhibited significant response recovery to the audible and to the combined audible/visible change but not to the visible change. In contrast, the 8-month-old infants exhibited significant response recovery to all three types of changes.

The results from the current experiment indicate that the 4- and 6-month-old infants discriminated only the A and A/V changes, whereas the 8-month-old infants discriminated all three types. This pattern of findings differs from the results reported by Lewkowicz (1996a) showing that infants responded to the V but not to the A changes when they were embedded in AD speech. The different outcome in the current experi-

Table 1 **Results of Planned Comparison Tests on the Data from the Test Trials in Experiment 1**

	Infant Age		
Test Trial	4 Months	6 Months	8 Months
Auditory	$F(1, 72) = 12.65$, $p < .001$	$F(1, 72) = 6.0$, $p < .025$	$F(1, 72) = 4.35$, $p < .05$
Visual	$F(1, 72) = .0$, $p = .98$	$F(1, 72) = 1.43$, $p = .24$	$F(1, 72) = 7.79$, $p < .01$
Auditory/Visual	$F(1, 72) = 5.56$, $p < .025$	$F(1, 72) = 5.39$, $p < .025$	$F(1, 72) = 7.79$, $p < .01$

ment may be due to one of two reasons. First, whereas in the current experiment infants saw the same person during the habituation and test phases, in Lewkowicz's (1996a) studies infants saw two different people during the two phases. Second, whereas in the current experiment the person in the habituation and the test phases was seen mouthing two different syllables, in Lewkowicz's (1996a) studies the two different people were seen mouthing two different passages. In other words, the visible differences in Lewkowicz's (1996a) studies were much more pronounced, and thus more discernible, than in the present experiment.

EXPERIMENT 2

The failure of the 4- and 6-month-old infants to detect the visible change in Experiment 1 may have resulted from the relative difficulty of detecting the visible contrast between two syllables. Therefore, in Experiment 2 the visible difference between the habituation phase and test phase syllables was increased by having the person speak in the AD manner during the habituation phase and in the ID manner during the test trial. One of the hallmarks of ID speech is that the person usually speaks at a slower pace, uses a higher pitch register, and modulates the voice over a greater frequency range (Fernald, 1985; Fernald & Kuhl, 1987). The visible correlates of this type of speech are that the person appears generally more animated and that she appears to be pronouncing each utterance in an exaggerated and slower manner. It was expected that the more pronounced visible attributes of ID speech would be more easily discriminable and that the 4- and 6-month-old infants would discriminate the visible change as well as the audible one.

Method

Participants. Separate groups of 4-, 6-, and 8-month-old infants ($n = 32$ per group) were tested, with an equal number of boys and girls in each group. The mean age of the 4-month group was 19.1 weeks ($SD = .92$ weeks). The mean age of the 6-month group was 27.7 weeks ($SD = .64$ weeks). One additional 6-month-old infant was tested but the data from this infant were not used because of inattentiveness. The mean age of the 8-month group was 36.7 weeks ($SD = .97$ weeks).

Apparatus, stimuli, and procedure. The apparatus and stimuli were the same as in Experiment 1, with one exception. In this experiment, the actor who was seen and heard during the test phase recited the syllables in the ID style. During the videotaping of the ID syllables, the actor was filmed in the presence of an awake 3-month-old infant (no infant sounds were heard on the recording) and was instructed to direct her speech toward the baby. In general, the actor smiled a good bit and was quite animated. Her speech also was exaggerated: Her voice was generally higher in pitch and more modulated than it was when she spoke the syllables in the AD manner. The average intensity of the syllables spoken during the habituation phase was the same as in Experiment 1 (71 dB for the first actor and 73 dB for the second one). The average intensity of the ID syllables spoken during the test trials was 72 dB for the first actor and 74 dB for the second one. In general, the variations in pitch and intensity were somewhat greater for the ID syllables (the intensity ranged between 3–4 dB). The testing procedure was the same as in Experiment 1. One 4-month-old, one 6-month-old, and four 8-month-old infants refused to sit alone and were tested on the parent's lap.

Results and Discussion

Figure 2 shows the results from the pretest, test, and posttest trials. A comparison of responsiveness during the pretest and posttest trials by way of a 2 (trials) × 3 (age) × 2 (gender) ANOVA indicated that the infants did not become fatigued during the test session. In addition, an analysis comparing the amount of looking in the posttest trial versus the amount of looking in the F test trial ($M = 7.8$ s) showed that infants recognized the novelty of the posttest stimulus, $F(1, 90) = 140.2$, $p < .001$.

As in Experiment 1, an initial analysis of responsiveness to the audible, visible, and audible/visible test changes was performed. It consisted of a mixed ANOVA with 3 (age), 2 (gender), and 4 (person × type of syllable combination) as between-subjects factors and 4 (trials) as a within-subjects factor. The analysis indicated that there was a significant effect of trials, $F(3, 216) = 46.68$, $p < .001$, a significant age effect, $F(2, 72) = 18.45$, $p < .001$. It also indicated that there was a significant Age × Trials interaction both without, $F(6, 216) = 2.61$, $p < .025$, and with the posttest trial data included, $F(8, 288) = 2.54$, $p < .025$.

As in Experiment 1, planned contrast analyses were performed at each age to determine whether the infants detected the test trial changes. As can be seen in Table 2, and as predicted, infants at all three ages exhibited response recovery to all three types of changes. This finding suggests that some aspect of multimodal ID speech made it easier for the infants in the current experiment to discriminate the visible change.

EXPERIMENT 3

The principal feature of the multimodal component variation method is that it permits an evaluation of

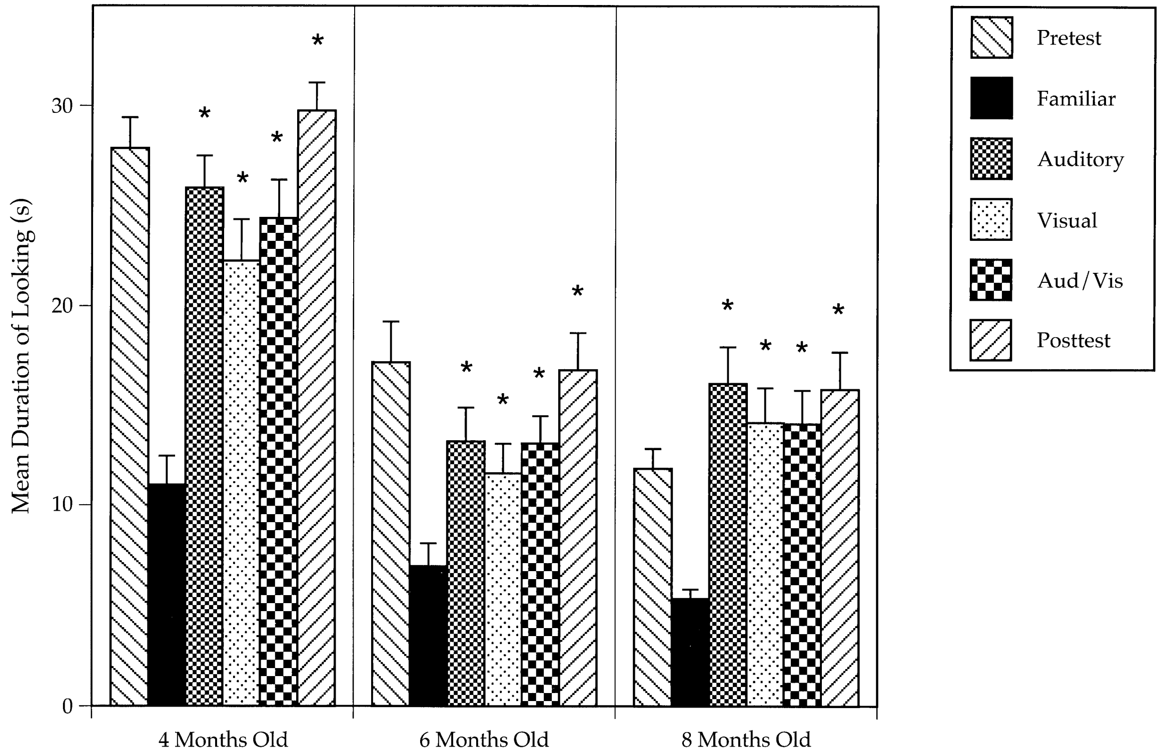

Figure 2 The results from Experiment 2 showing the mean duration of looking plotted separately for each age group during the pretest, test, and posttest trials. Error bars signify standard error of the mean and asterisks indicate significant response recovery.

responsiveness to the components making up a multimodal compound stimulus. This is done by changing some aspect of the information in one sensory modality while leaving the information in the other modality unchanged. In the present experiments, such changes led to the concomitant disruption of the temporal relationship between the auditory and visual components. Therefore, it cannot be determined whether infants were responding to the temporal discrepancy in the A and V test trials or to the differences in one component across the habituation and test phases.

At a minimum, the data from the previous two experiments suggest that responsiveness to the temporal discrepancy could not have been the principal reason for the observed findings. First, neither the 4- nor the 6-month-old infants in Experiment 1 responded to the change in the V test trial even though the synchrony between the facial and vocal information in that test trial was disrupted. Second, infants in both experiments responded to the bimodal change in the A/V test trial despite the fact that during this test trial there was no desynchronization of the A and V components. Finally, Lewkowicz (1996a, 1998) obtained a similar pattern of findings in the continuous utterance studies, although the asynchrony in those studies changed continuously during the course of a test trial. Despite these findings, however, neither the continuous utterance studies nor Experiments 1 and 2 provide direct evidence on whether infants perceive AV temporal synchrony in speech. Work with adults

Table 2 Results of Planned Comparison Tests on the Data from the Test Trials in Experiment 2

	Infant Age		
Test Trial	4 Month	6 Month	8 Month
Auditory	$F(1, 72) = 67.38, p < .001$	$F(1, 72) = 11.85, p < .001$	$F(1, 72) = 35.17, p < .001$
Visual	$F(1, 72) = 42.06, p < .001$	$F(1, 72) = 7.18, p < .01$	$F(1, 72) = 25.73, p < .001$
Auditory/Visual	$F(1, 72) = 72.45, p < .001$	$F(1, 72) = 15.26, p < .001$	$F(1, 72) = 30.89, p < .001$

(Dixon & Spitz, 1980; Massaro, 1998; McGrath & Summerfield, 1985; Radeau, 1994) and infants (Bahrick, 1992; Lewkowicz, 1992a, 1996b; Spelke, Born, & Chu, 1983; Walker, 1982) has shown that temporal AV relations contribute in important ways to intersensory integration. Moreover, Dodd (1979) has reported that infants prefer to look more at bimodal continuous speech for which the audible and visible components are in synchrony than at the same speech for which the components are desynchronized. Thus, it is likely that synchrony relations may be important to multimodal speech perception, at least in a supportive sense. This conclusion is supported by Walker's (1982) findings that infants can detect the A and V attributes of affective expressions even when they are asynchronous but that they also are sensitive to the temporal relations between the two attributes.

To determine whether infants are sensitive to AV temporal synchrony relations at the syllabic level, infants were habituated to a talking face whose vocal and facial features were synchronous and then were given two types of test trials. The first test trial was designed to test responsiveness to asynchrony alone and thus involved the presentation of a talking face whose vocal and facial components were asynchronous. Choosing a specific degree of asynchrony was problematic because no prior empirical studies have investigated infants' responsiveness to syllabic AV asynchrony. Studies of infants' response to talking faces (Lewkowicz, 1996a, 1998) have shown that infants do not respond to desynchronized audible and visible speech in continuous speech. On the other hand, studies of infants' responsiveness to AV asynchrony in a simple AV compound stimulus (i.e., a bouncing object and a tone) have shown that they can detect an asynchrony above 350 ms (Lewkowicz, 1996b). Therefore, an asynchrony of 666 ms was selected because it is greater than that detected in simpler events.

The second test trial also involved the presentation of asynchronous vocal and facial components, but in addition, these components were produced by a novel person. This was done because infants often have to attend to more than one person talking at the same time and thus must be able to perceive the temporal relation of the A and V components specifying multiple people. In other words, their perception of AV temporal relations has to occur in the context of competing featural information. In fact, this was the task that infants in Lewkowicz's (1996a, 1998) continuous utterance studies were asked to perform; however, Lewkowicz did not specifically assess the role of featural information in responsiveness. Given the general importance of this issue, the current experiment was designed, in part, to address it.

Method

Participants. Separate groups of 4-, 6-, and 8-month-old infants ($n = 32$ per group) were tested, with an equal number of boys and girls in each group. The mean age of the 4-month group was 18.7 weeks ($SD =$ .9 weeks). Five additional 4-month-old infants were tested but their data were not used because four of them fussed and one was sleepy. The mean age of the 6-month group was 27.9 weeks ($SD = 1.7$ weeks). Five additional 6-month-old infants were tested but the data from these infants were not used because they were fussy. The mean age of the 8-month group was 36.5 weeks ($SD = 1$ week). Two additional 8-month-old infants were tested but their data were not used because they were fussy.

Apparatus, stimuli, and procedure. The apparatus and procedure were the same as in Experiment 1 except that the stimuli differed in the following way. Because analyses of the data from Experiments 1 and 2 indicated that the results were not affected by the specific syllable used, all infants in this experiment were habituated to the syllable /ba/. Half the infants at each age were habituated to this syllable spoken by one of the actors presented in the prior two experiments and the other half of the infants were habituated with this syllable spoken by the other actor. To determine whether the specific manner of speech might play a role in responsiveness, each of these groups was further divided in half, with one habituated and tested with the /ba/ syllable spoken in the ID manner and the other half habituated and tested with the same syllable spoken in the AD manner. As a result, there were four habituation groups of eight infants each at each of the three ages.

Following habituation, each infant was given a series of three test trials in a fixed order. The first was an F test trial during which the infant saw and heard the familiar person with synchronous vocal and facial attributes. The second was an asynchrony/familiar person (asncFP) test trial where the infant saw and heard the familiar person except that the person's voice preceded the motion of her lips by 666 ms (20 video frames). The third test trial was an asynchrony/novel person (asncNP) test trial where the infant saw and heard a novel person (the other actor) whose voice also preceded the visible motion of her lips by 666 ms. As in the prior experiments, each infant also received the pre- and posttest trials during which the cartoon segment was played. The average intensity of the soundtrack of one of the female actors was 71 dB while reciting the AD syllable and 73 dB while reciting the ID syllable, with slightly greater variation in intensity as well as in pitch for the latter. The average intensity of the soundtrack of the other female was 74 dB while reciting the AD syllable and 75 dB while reciting the ID syllable, also with slightly greater variation in intensity and in pitch for the latter. Seven 4-month-old, seven 6-month-old, and four 8-month-old infants refused to sit alone and were tested on the parent's lap.

Results and Discussion

The results from the pretest, test, and posttest trials can be seen in Figure 3. A comparison of responsiveness in the pretest and posttest trials by way of a 2 (trials) × 3 (age) × 2 (gender) ANOVA indicated that infants did not become fatigued during the test session in that there was no difference in looking in the pretest ($M = 18$ s) and the posttest ($M = 18.2$ s) trials. In addition, an analysis comparing the amount of looking in the posttest trial versus the amount of looking in the F test trial ($M = 8.3$ s) showed that infants recognized the novelty of the posttest stimulus, $F(1, 90) = 76.3, p < .001$.

To determine whether infants discriminated the change in the temporal relationship alone or the temporal relationship + person change, the data from the F, asncFP, and asncNP test trials were submitted to a mixed ANOVA, with 3 (age), 2 (gender), and 4 (person × manner-of-speech combination) as between-subjects factors and 3 (trials) as a within-subjects factor. Results indicated that there was a significant effect of trials, $F(2, 144) = 35.98, p < .001$, and of age, $F(2, 72) = 3.52, p < .05$, and that the Age × Trials interaction was marginally significant, $F(4, 144) = 2.35$, $p < .06$. When the posttest trial data were included in the analysis, the Age × Trials interaction was significant, $F(6, 216) = 5.51, p < .001$. The Person × Manner of speech variable did not exert a significant effect ($p = .90$), nor did it interact with trials ($p > .30$).

Planned comparison tests were performed separately at each age comparing the duration of looking in the F test trial versus the duration of looking in the two asynchrony test trials, respectively. As can be seen in Table 3, the 4-month-old infants did not exhibit response recovery in the asncFP test trial but did in the asncNP test trial. In contrast, the two older groups of infants exhibited response recovery to both types of changes. These results suggest that the

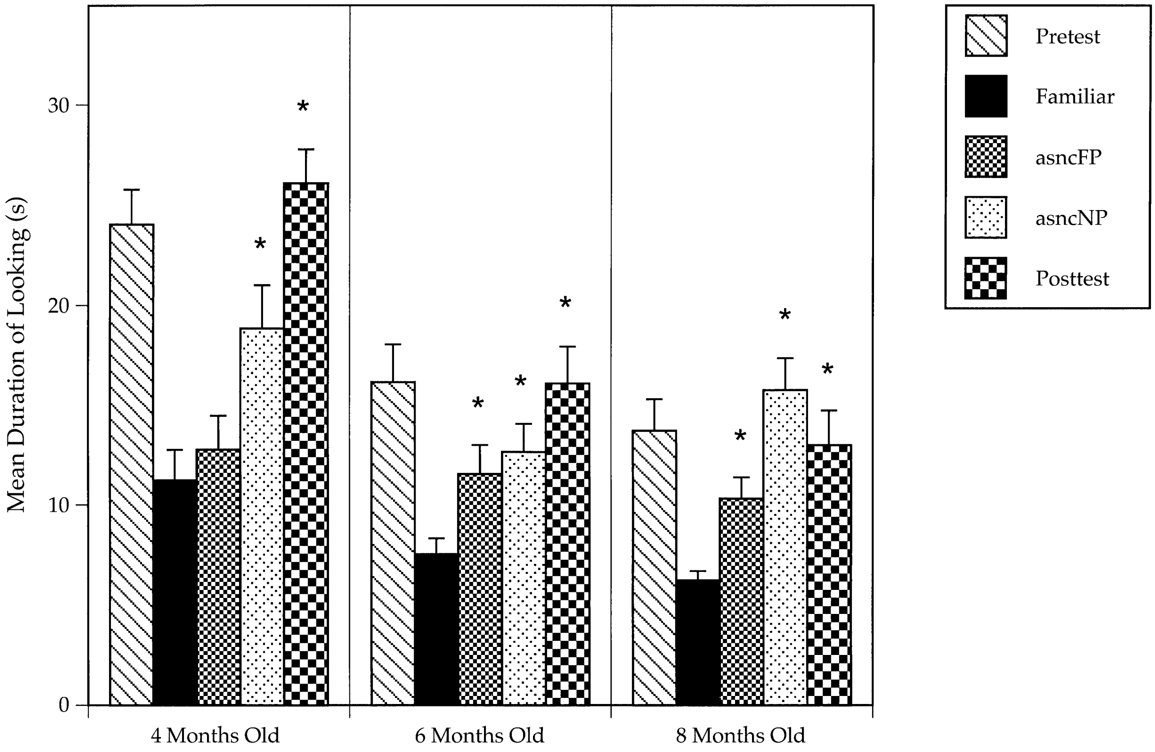

Figure 3 The results from Experiment 3 showing the mean duration of looking during the pretest, familiar, asynchronous familiar-person (asncFP), asynchronous novel-person (asncAP), and posttest trials plotted separately for each age group. Error bars signify standard error of the mean and asterisks indicate significant response recovery.

Table 3 Results of Planned Comparison Tests on the Data from the Test Trials in Experiment 3

	Infant Age		
Test Trial	4 Months	6 Months	8 Months
asncFP	$F(1, 72) = 1.16, p = .28$	$F(1, 72) = 7.98, p < .01$	$F(1, 72) = 7.1, p < .01$
asncNP	$F(1, 72) = 18.94, p < .001$	$F(1, 72) = 8.66, p < .01$	$F(1, 72) = 31.51, p < .001$

Note: asncFP = asychrony/familiar person; asncNP = asychronyl/novel person.

4-month-old infants did not perceive the change in the temporal relation between the vocal and facial attributes of the talking face but that the two older groups of infants did. Furthermore, the results suggest that the 4-month-old infants' discrimination in the asncNP test trial was due to the detection of the change in the vocal, facial, or both components of the person but not to the disruption of their temporal contiguity.

To determine whether the concurrent availability of the bimodal featural information enhanced discriminability, Scheffé post hoc tests comparing the duration of looking in the asncFP and asncNP test trials were performed. The results of these tests indicated that the 4-month-old infants looked more to the asncNP than to the asncFP change, $p < .001$, as did the 8-month-old infants, $p < .001$. The same analyses showed that the 6-month-old infants increased their looking time to each type of change to the same degree.

When considered together, the planned contrast and post hoc analyses suggest an interesting developmental pattern. In the 4-month-old infants the absence of increased looking to asynchrony in the asncFP test trial coupled with increased looking to the featural change in the asncNP test trial suggests that the increased looking in the asncNP test trial was due either to the detection of the featural change alone or to the detection of the concurrent temporal/featural change. One cannot, however, distinguish between these two alternative explanations without introducing a test trial in which only a feature is changed. The pattern of findings in the 6-month-old infants suggests that they based their responsiveness on temporal synchrony relations in both test trials. This conclusion stems from the equivalent amount of looking in the asncFP and asncNP trials. Finally, the findings from the 8-month-old infants suggest that they detected the synchrony relation as well as the featural changes because they exhibited greater looking in the asncNP test trial.

The absence of manner-of-speech effects is interesting. It suggests that, when it emerges, responsiveness to intersensory temporal synchrony relations in speech is relatively robust in that its perception is not diminished when the audible prosody of the syllable and its visible correlates is minimized. It might be hypothesized, however, that if the ability to perceive synchrony relations between the vocal and facial attributes of a talking face is in the process of emerging (i.e., around 4 months of age) then its perception might be more susceptible to manner-of-speech effects. To determine whether there was any indication of such susceptibility, a separate analysis of the data from just those 4-month-old infants who were exposed to the ID syllables was conducted. This analysis showed that these infants did not exhibit recovery in the asncFP test trial, $F(1, 15) = 1.43$, $p = .25$, nor in the asncNP test trial, $F(1, 15) = 3.74$, $p = .07$.

GENERAL DISCUSSION

Three experiments were conducted to investigate infants' perception of the audible, visible, and combined components of isolated syllables. The results from Experiment 1 indicated that following habituation to a talking face uttering either the syllable /ba/ or /sha/ in the AD manner, 4- and 6-month-old infants responded only to the audible and bimodal changes and not to the visible ones. In contrast, 8-month-old infants responded to all three types of changes. Results from Experiment 2 showed that when the difference between the habituation and test phase was enhanced (by habituating infants to a person speaking in the AD manner and testing them with a person speaking in the ID manner), all age groups discriminated all three types of changes. Because the unimodal test trials produced a desynchrony of the audible and visible attributes of the syllable, Experiment 3 tested the role that temporal synchrony might have played in responsiveness. In addition, this experiment examined the role of concurrent featural cues in discrimination. Results suggested that responsiveness to concurrent temporal synchrony and featural cues changes over development in that at 4 months of age infants attended only to the featural cues, at 6 months only to temporal synchrony cues, and at 8

Table 4 Summary of the Results from Lewkowicz's (1996a, 1998) Studies of Infants' Response to Bimodal Continuous Utterances and from Experiments 1 and 2

	Continuous Utterance Studies				Single Syllable Studies		
Type of Change	Auditory	Visual	Auditory/Visual	Type of Change	Auditory	Visual	Auditory/Visual
AD→AD				AD→AD			
4 months	—	—	—	4 months	√	—	√
6 months	—	√	√	6 months	√	—	√
8 months	—	√	√	8 months	√	√	√
AD→AD & Gender				AD→ID			
4 months	—	√	√	4 months	√	√	√
6 months	—	√	√	6 months	√	√	√
8 months	—	√	√	8 months	√	√	√
AD→ID & Gender							
4 months	—	√	√				
6 months	√	√	√				
8 months	√	√	√				
AD→Song							
3 months	—	√	√				
4 months	√	√	√				
6 months	√	√	√				
8 months	√	√	√				
AD→Song & Gender							
3 months	√	√	√				
4 months	√	√	√				
6 months	√	√	√				
8 months	√	√	√				

Note: √ means a successful discrimination and − means failure to respond.

months to both types of cues. Taken together, the current results point to some of the dimensions of multimodal speech that are important during early development.

Kinematic Information

Table 4 presents a summary of both the findings from the previous studies from this laboratory on infants' response to continuous utterances (Lewkowicz, 1988a, 1988b) and the findings from Experiments 1 and 2 on infants' response to single syllables. As can be seen in Table 4, with the exception of the 4-month-old infants in the AD→AD condition, all the infants in the continuous utterance studies responded to the visible change. In contrast, the two youngest groups of infants in Experiment 1 from the current study did not respond to the visible change. This finding is similar to findings from studies presenting flashing checkerboards and pulsing tones (Lewkowicz, 1988a, 1988b). In these studies, infants were habituated to inanimate, spatially static AV compound stimuli consisting of a flashing checkerboard and a pulsing tone and then tested with the multimodal component variation method for responsiveness to A, V, and AV changes in rate of presentation. Results indicated that 6-month-old infants did not respond to visual component changes but did respond to auditory component changes, whereas 10-month-old infants responded, although less so, to visual component changes as well as to auditory component changes. In other words, when kinematic variations were absent, infants responded more to the auditory component changes. The similarity of the current findings to those from the studies using flashing checkerboards is interesting because it suggests that the reason for the infants' failure in Experiment 1 to respond to the visual differences may be the reduced amount of kinematic visual information that was available. According to this logic, infants responded to the visible changes in the continuous utterance studies was because the amount of motion available was maximal: A person could be seen either speaking a continuously variable script or singing and, as a result, the person projected a good deal of kinematic visual information. This contrasts with the case of the single syllable, where far less kinematic information is available.

The findings from the talking face studies, on the one hand, and the findings from the checkerboard/sound studies, on the other, appear to represent two ends of a continuum whose critical feature with regard to visual responsiveness appears to be motion. Indeed, findings from another set of studies (Lewkowicz, 1992b) in which infants' response to an A/V compound stimulus for which the visual component was spatially dynamic are consistent with the motion-based interpretation. In these studies, infants were habituated to a simple, two-dimensional visual stimulus that bounced and made a sound each time it bounced. Infants as young as four months of age had no difficulty discriminating changes in the visual component and, in fact, exhibited a more robust discrimination of the visible component change than the audible one.

Needless to say, there are different types and degrees of motion and it is clear that this plays an important role in infant responsiveness (Dannemiller & Freedland, 1991; Roessler & Dannemiller, 1997). For example, in terms of the overall degree of motion, a person who is engaged in the act of singing typically projects more motion than a person engaged in the act of speaking. Moreover, when speaking, the degree of motion produced by a person's face, lips, and head can depend on the specific type of speech that the person produces. Perhaps the least dynamic is the face and voice of a person reciting a single syllable, and even then, the degree and type of motion that is visible can differ depending on the specific type of syllable produced. For example, the syllable /ba/ is characterized by clearly visible, well-defined lip closure at its onset and is then followed by lip and mouth opening. In contrast, the syllable /sha/ is characterized primarily by open lips.

At a basic level, the perception of a talking face depends on the general process of object perception. Kellman and Arterberry (1998) have summarized the research on infants' object perception abilities and have concluded that object perception undergoes a major shift around the 7th month of life. Initially, infant object perception is based on what Kellman and Arterberry term an edge-insensitive process. This process is responsive primarily to the kinematic properties of objects and is insensitive to edge relationships. Indeed, infants are highly sensitive to visual motion right from birth (Dannemiller & Freedland, 1991; Kaufmann, Stucki, & Kaufmann-Hayoz, 1985; Volkmann & Dobson, 1976) and are more responsive to moving visual displays than to spatially-static ones (Volkmann & Dobson, 1976; Wilcox & Clayton, 1968). According to Kellman and Arterberry (1998), the edge-insensitive process is replaced by an edge-sensitive process by 7 months of age. The edge-sensitive process provides infants with a much richer view of the world because it not only specifies the kinematic properties of objects and events but also specifies object connectedness and the forms of hidden boundaries. Thus, by 7 months of age infants become capable of using both the kinematic properties of objects and their edge information allowing for perception of the objects in a more detailed and more accurate fashion.

I propose that this kind of shift, together with the general process of perceptual differentiation (Gibson, 1969; Werner, 1973), most likely accounts for the age difference in infants' response to the visible syllable changes found in Experiment 1. That is, although infants attended to kinematic differences, the differences between the two syllables were probably not sufficiently distinct for the 4- and 6-month-old infants. Moreover, these infants probably did not detect the finer visual aspects of the two syllables (i.e., the difference between open and closed lips) because the edge-sensitive mechanism had not developed yet. In contrast, the 8-month-old infants were able to rely on edge-sensitive mechanisms to detect the finer details and so perceptually differentiated the two syllables.

Sensory Salience Hierarchies

An additional factor that might have affected infants' detection of the visible and audible features of the syllables in Experiment 1 is the differential role that visual and auditory information plays in specifying speech. Evidence from adult studies indicates that even though visual information contributes in important ways to the perception of speech (Massaro, 1998), it plays a weaker role than does auditory information in the identification of syllables. For example, Massaro, Cohen, and Smeele (1996) studied responsiveness to different degrees of asynchrony between concurrently presented visible and audible syllables. On some trials the audible and visible syllables were the same but on others they were different. When the audible and visible syllables were different, identification was disrupted but the disruption was greater when participants had to perform visual identification with inconsistent auditory information than it was when participants had to make auditory identification with inconsistent visual information. In other words, auditory information played a greater role in syllable identification than did visual information. This finding suggests that the poorer power of visual information to specify the identity of syllables together with poorer object perception abilities in infants younger than 7 months of age probably prevented the two younger groups of infants from detecting changes in the visible aspects of the syllables.

The finding that the 4- and 6-month-old infants in Experiment 2 successfully discriminated the visible syllable changes is consistent with the developmental interpretation offered above. Infant-directed speech is generally more salient than AD speech in that the audible attributes of ID speech are characterized by highly modulated frequency contours, a slower overall tempo, and exaggerated temporal structure consisting of elongated pauses. With specific regard to the visible attributes of infant-directed speech, a person's facial and head movements are usually more dynamic. Doubtless, these types of exaggerated features are highly salient and thus are highly attention getting. Indeed, research has shown that infants as young as 1 month of age prefer auditory ID over AD speech (Cooper & Aslin, 1990; Fernald, 1985; Werker & McLeod, 1989) and that infants as young as 4 months of age prefer auditory/visual ID speech (Werker & McLeod, 1989). Thus, the greater kinematic range of the ID syllables that created greater visible differences most likely contributed to the successful discrimination of the visible component changes in Experiment 2 in the two youngest groups of infants.

Audible Differences

As can be seen in Table 4, when 4-, 6-, and 8-month-old infants were tested for continuous utterance discrimination between two different people either of the same gender or of different genders speaking in the AD manner (Experiments 1 and 2 in Lewkowicz, 1996a), they did not discriminate the audible differences. When these same age groups were tested for continuous utterance discrimination between a male person speaking in the AD manner and a female person speaking in the ID manner (Experiment 3 in Lewkowicz, 1996a), the two older groups discriminated the audible differences but the 4-month-old infants still did not. Finally, whereas 4-, 6-, and 8-month-old infants discriminated the audible difference between a person speaking in the AD manner and a person of the same gender who was singing (Experiment 1 in Lewkowicz, 1998), 3-month-old infants did not discriminate this difference. When, however, the person singing was of a different gender the 3-month-olds, as well as the three older groups, discriminated the audible difference. In the case of the current experiments, the data are consistent with the initial suggestion that the complexity of the auditory information affects the way infants respond to aspects of audible speech in a multimodal context. In contrast to the difficulty that infants in the continuous utterance studies had in responding to the audible differences, the infants in the current experiments had no difficulty detecting the audible differences when they were inherent in single-syllable contrasts.

Temporal Relations

Although the findings from Experiment 1 indicated that asynchrony could not have been the sole determinant of responsiveness in the unimodal test trials, the findings from the first two experiments raised the possibility that, when available, temporal synchrony relations might have contributed in some way. The results from Experiment 3 indicated that the 4-month-old infants did not respond to the AV asynchrony despite its size (666 ms). This finding appears to be consistent with the findings on infants' response to AV asynchronies in continuous utterances (Lewkowicz, 1996a) in that infants between 4 and 8 months of age did not respond to the asynchrony. It is impossible to know, however, what the specific degree of asynchrony was in those studies because it varied continuously throughout the test trial. A more appropriate comparison would be data from studies in which the AV asynchrony was kept at a constant level. Dodd's (1979) study is one in which continuous speech, with a constant AV asynchrony of 400 ms, was presented; infants as young as 3 months of age discriminated synchrony from asynchrony by showing a preference for synchronous speech. Likewise, studies by Lewkowicz (1986, 1996b) in which the synchrony between the audible and visible attributes of nonspeech events was disrupted indicate that infants as young as 4 months of age detect AV asynchronies in the region between 300 and 400 ms.

What might be the reason for the failure of the 4-month-old infants in Experiment 1 to perceive the AV asynchrony? Earlier it was argued that the developmental shift from edge-insensitive to edge-sensitive mechanisms makes it possible for infants to detect the finer visible aspects of the individual syllables and that this makes it possible for infants older than 7 months to detect the visible attributes of the talking face. The shift described by Kellman and Arterberry (1998), however, applies specifically to visual perception. It may be that responsiveness to the relationship between visual information on the one hand, and auditory information on the other, might follow a somewhat different developmental schedule. In fact, one of the most consistent and reliable findings in the literature on intersensory perception in both human and animal infants has been the fact that responsiveness to bimodal stimulation is far greater than is responsiveness to unimodal stimulation (Lewkowicz, 1988a, 1988b, 1996a, 1998; Stein & Meredith, 1993; Stein, Meredith, & Wallace, 1994). Given this sensitivity, it may be that infants can use the edge-sensitive mechanism by 6 months of age when they attempt to relate the visual and auditory information in terms of its temporal contiguity.

Methodological Considerations

I have argued that the differences between the results from Lewkowicz's (1996a, 1998) studies and those from the present experiments reflect differences in the complexity of the multimodal utterances. It is possible, however, that the differences are due, in part, to the fact that an infant-controlled habituation procedure was used in the current studies whereas a fixed-trials habituation procedure was used by Lewkowicz (1996a, 1998). Because of the procedural differences, it might be argued that different amounts of overall exposure to the habituation stimulus might have allowed infants to differentially encode it and that this, in turn, might have led to differences in responsiveness during the test trials. Indeed, a comparison of the total amount of looking during the habituation phase in Experiments 1, 2, and 3 ($M = 138$ s) in Lewkowicz (1996a) versus the total amount of looking in the current experiments ($M = 113$ s) indicated that the infants in the former studies spent more time looking, $F(1, 570) = 35.75, p < .001$. This makes it possible that the differences in outcome may be partly ascribed to the procedural differences. On the other hand, evidence indicates that even very short (e.g., 10 s) familiarization procedures can produce a novelty response in infants (Bornstein, 1985). This suggests that the relatively extended amounts of exposure that infants received in the two sets of studies were more than sufficient to produce a discriminative response and that the different outcomes were not due to the differential sensitivity of the two procedures. The ultimate answer to this question will have to await experiments that directly compare responsiveness to the same stimulus contrasts across the two types of habituation procedures.

Conclusions

The current findings add to our knowledge about infants' perception of faces and voices. When they are considered together with previous findings, they suggest that the complexity of bimodal speech plays a role in responsiveness. It appears that infants are more likely to discriminate the audible features of bimodal speech when the speech consists of an isolated syllable and that the emergence of responsiveness to the audible attributes of bimodal syllables precedes the emergence of responsiveness to its visible attributes. The current findings also show that the temporal relationship between the audible and visible attributes of syllables becomes a salient perceptual characteristic by 6 months of age and that featural characteristics associated with a bimodal syllable become salient by 8 months of age. Taken together, these findings provide interesting insights into infants' perception of multimodal events and the perceptual mechanisms that may contribute to the development of person perception and general communicative skills.

ACKNOWLEDGMENTS

This work was supported in part by funds from the New York State Office of Mental Retardation and Developmental Disabilities and by grant RO3 HD36731 from NICHD. The author thanks Marci Dabbene for her assistance and the staff at St. Vincent's Medical Center of Richmond County, Staten Island, New York, for their assistance in subject recruitment. He also thanks Leslie Cohen for his generosity in providing the software used to run these studies and Linnea Dickson for comments on an earlier draft.

ADDRESS AND AFFILIATION

Corresponding author: David J. Lewkowicz, New York State Institute for Basic Research in Developmental Disabilities, 1050 Forest Hill Rd., Staten Island, NY 10314; e-mail: ddlsi@cunyvm.cuny.edu.

REFERENCES

Bahrick, L. E. (1992). Infants' perceptual differentiation of amodal and modality-specific audio-visual relations. *Journal of Experimental Child Psychology, 53*, 180–199.

Bahrick, L. E. (1994). The development of infants' sensitivity to arbitrary intermodal relations. *Ecological Psychology, 6*, 111–123.

Bahrick, L. E. (1996, April). *Increasing specificity in the development of infants' sensitivity to two nested amodal relations.* Paper presented at the International Conference on Infant Studies, Providence, RI.

Bertelson, P., & Radeau, M. (1976). Ventriloquism, sensory interaction, and response bias: Remarks on the paper by Choe, Welch, Gilford, and Juola. *Perception & Psychophysics, 19*, 531–535.

Bertelson, P., & Radeau, M. (1981). Cross-modal bias and perceptual fusion with auditory-visual spatial discordance. *Perception & Psychophysics, 29*, 578–584.

Birch, H. G., & Lefford, A. (1963). Intersensory development in children. *Monographs of the Society for Research in Child Development, 25*(Serial No. 89).

Birch, H. G., & Lefford, A. (1967). Visual differentiation, intersensory integration, and voluntary motor control. *Monographs of the Society for Research in Child Development, 32* (Serial No. 110).

Bornstein, M. H. (1985). Habituation of attention as a measure of visual information processing in human infants: Summary, systematization, and synthesis. In G. Gottlieb & N. A. Krasnegor (Eds.), *Measurement of audition and vision in the first year of postnatal life: A methodological overview* (pp. 253–300). Norwood, NJ: Ablex.

Burnham, D. (1998). Language specificity in the development of auditory-visual speech perception. In R. Campbell, B. Dodd, & D. Burnham (Eds.), *Hearing by eye II: Advances in the psychology of speechreading and auditory-visual speech* (pp. 27–60). Hove, U.K.: Psychology Press/Erlbaum (U.K.) Taylor & Francis.

Cooper, R. P., & Aslin, R. N. (1990). Preference for infant-directed speech in the first month after birth. *Child Development, 61*, 1584–1595.

Dannemiller, J. L., & Freedland, R. L. (1991). Detection of relative motion by human infants. *Developmental Psychology, 27*, 67–78.

Dixon, N. F., & Spitz, L. T. (1980). The detection of auditory visual desynchrony. *Perception, 9*, 719–721.

Dodd, B. (1979). Lip reading in infants: Attention to speech presented in- and out-of-synchrony. *Cognitive Psychology, 11*, 478–484.

Edelman, G. M. (1992). *Bright air, brilliant fire.* New York: Basic Books.

Fernald, A. (1985). Five-month-old infants prefer to listen to motherese. *Infant Behavior and Development, 8*, 181–195.

Fernald, A., & Kuhl, P. K. (1987). Acoustic determinants of infant preference for motherese. *Infant Behavior and Development, 10*, 279–293.

Gibson, E. J. (1969). *Principles of perceptual learning and development.* New York: Appleton.

Gibson, E. J. (1984). Perceptual development from the ecological approach. In M. E. Lamb, A. L. Brown, & B. Rogoff (Eds.), *Advances in Developmental Psychology* (pp. 243–286). Hillsdale, NJ: Erlbaum.

Gibson, J. J. (1966). *The senses considered as perceptual systems.* Boston: Houghton-Mifflin.

Gottlieb, G. (1971). Ontogenesis of sensory function in birds and mammals. In E. Tobach, L. R. Aronson, & E. Shaw (Eds.), *The biopsychology of development* (pp. 67–128). New York: Academic Press.

Horowitz, F. D. (1974). Visual attention, auditory stimulation, and language discrimination in young infants. *Monographs of the Society for Research in Child Development, 39*(Serial No. 158).

Jack, C. E., & Thurlow, W. R. (1973). Effects of degree of visual association and angle of displacement on the "ventriloquism" effect. *Perceptual and Motor Skills, 37*, 967–979.

Jusczyk, P. W., Houston, D., & Goodman, M. (1998). Speech perception during the first year. In A. Slater (Ed.), *Perceptual development: Visual, auditory, and speech perception in infancy* (pp. 357–388). East Sussex, U.K.: Psychology Press.

Jusczyk, P. W., & Kemler Nelson, D. G. (1996). Syntactic units, prosody, and psychological reality during infancy. In J. L. Morgan & K. Demuth (Eds.), *Signal to syntax: Bootstrapping from speech to grammar in early acquisition* (pp. 389–408). Mahwah, NJ: Erlbaum.

Kaufmann, F., Stucki, M., & Kaufmann-Hayoz, R. (1985). Development of infants' sensitivity for slow and rapid motions. *Infant Behavior and Development, 8*, 89–98.

Kellman, P. J., & Arterberry, M. E. (1998). *The cradle of knowledge: Development of perception in infancy.* Cambridge, MA: MIT Press.

Kuhl, P. K., & Meltzoff, A. N. (1982). The bimodal perception of speech in infancy. *Science, 218*, 1138–1140.

Lewkowicz, D. J. (1986). Developmental changes in infants' bisensory response to synchronous durations. *Infant Behavior & Development, 9*(3), 335–353.

Lewkowicz, D. J. (1988a). Sensory dominance in infants: I. Six-month-old infants' response to auditory-visual compounds. *Developmental Psychology, 24*, 155–171.

Lewkowicz, D. J. (1988b). Sensory dominance in infants: II. Ten-month-old infants' response to auditory-visual compounds. *Developmental Psychology, 24*, 172–182.

Lewkowicz, D. J. (1992a). Infants' response to temporally based intersensory equivalence: The effect of synchronous sounds on visual preferences for moving stimuli. *Infant Behavior & Development, 15*, 297–324.

Lewkowicz, D. J. (1992b). Infants' responsiveness to the auditory and visual attributes of a sounding/moving stimulus. *Perception & Psychophysics, 52*, 519–528.

Lewkowicz, D. J. (1996a). Infants' response to the audible and visible properties of the human face. I: Role of lexical-syntactic content, temporal synchrony, gender, and manner of speech. *Developmental Psychology, 32*, 347–366.

Lewkowicz, D. J. (1996b). Perception of auditory-visual temporal synchrony in human infants. *Journal of Experimental Psychology: Human Perception & Performance, 22*, 1094–1106.

Lewkowicz, D. J. (1998). Infants' response to the audible and visible properties of the human face: II. Discrimination of differences between singing and adult-directed speech. *Developmental Psychobiology, 32*, 261–274.

Lewkowicz, D. J. (1999). The development of temporal and spatial intermodal perception. In G. Aschersleben, T. Bachman, & J. Müsseler (Eds.), *Cognitive contributions to the perception of spatial and temporal events* (pp. 395–420). Amsterdam: Elsevier.

Lewkowicz, D. J., & Lickliter, R. (Eds.). (1994). *Development of intersensory perception: Comparative perspectives.* Hillsdale, NJ: Erlbaum.

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

Maier, N. R. F., & Schneirla, T. C. (1964). *Principles of animal psychology.* New York: Dover.

Marks, L. (1978). *The unity of the senses.* New York: Academic Press.

Massaro, D. W. (1998). *Perceiving talking faces: From speech perception to a behavioral principle.* Cambridge, MA: MIT Press.

Massaro, D. W., Cohen, M. M., & Smeele, P. M. T. (1996). Perception of asynchronous and conflicting visual and auditory speech. *Journal of the Acoustical Society of America, 100*, 1777–1786.

McGrath, M., & Summerfield, Q. (1985). Intermodal timing relations and audio-visual speech recognition by normal-hearing adults. *Journal of the Acoustical Society of America, 77*, 678–685.

McGurk, H., & MacDonald, J. (1976). Hearing lips and seeing voices. *Nature, 264*, 229–239.

Miller, C. L. (1983). Developmental changes in male / female voice classification by infants. *Infant Behavior and Development, 6*, 313–330.

Piaget, J. (1954). *The construction of reality in the child.* London: Routledge and Kegan Paul.

Radeau, M. (1994). Auditory-visual spatial interaction and modularity. *Cahiers de Psychologie, 13*, 3–51.

Roessler, J. S., & Dannemiller, J. L. (1997). Changes in human infants' sensitivity to slow displacements over the first 6 months. *Vision Research, 37*, 417–423.

Rosenblum, L. D., Schmuckler, M. A., & Johnson, J. A. (1997). The McGurk effect in infants. *Perception & Psychophysics, 59*, 347–357.

Spelke, E. S., Born, W. S., & Chu, F. (1983). Perception of moving, sounding objects by 4-month-old infants. *Perception, 12*, 719–732.

Stein, B. E., & Meredith, M. A. (1993). *The merging of the senses.* Cambridge, MA: MIT Press.

Stein, B. E., Meredith, M. A., & Wallace, M. T. (1994). Development and neural basis of multisensory integration. In D. J. Lewkowicz & R. Lickliter (Eds.), *The development of intersensory perception: Comparative perspectives* (pp. 81–105). Hillsdale, NJ: Erlbaum.

Thelen, E., & Smith, L. B. (1994). *A dynamic systems approach to the development of cognition and action.* Cambridge, MA: MIT Press.

Volkmann, F. C., & Dobson, M. V. (1976). Infant responses of ocular fixation to moving visual stimuli. *Journal of Experimental Child Psychology, 22*, 86–99.

Walker, A. S. (1982). Intermodal perception of expressive behaviors by human infants. *Journal of Experimental Child Psychology, 33*, 514–535.

Walker-Andrews, A. S. (1986). Intermodal perception of expressive behaviors: Relation of eye and voice? *Developmental Psychology, 22*, 373–377.

Walker-Andrews, A. S. (1997). Infants' perception of expressive behaviors: Differentiation of multimodal information. *Psychological Bulletin, 121*, 437–456.

Welch, R. B., & Warren, D. H. (1986). Intersensory interactions. In K. R. Boff, L. Kaufman, & J. P. Thomas (Eds.), *Handbook of perception and human performance: Sensory processes and perception* (Vol. 1, pp. 1–36). New York: Wiley.

Werker, J. F., & McLeod, P. J. (1989). Infant preference for both male and female infant-directed talk: A developmental study of attentional and affective responsiveness. *Canadian Journal Psychology, 43*, 230–246.

Werner, H. (1973). *Comparative psychology of mental development.* New York: International Universities Press.

Wilcox, B. M., & Clayton, F. L. (1968). Infant visual fixation on motion pictures of the human face. *Journal of Experimental Child Psychology, 6*, 22–32.