INFANTS' PERCEPTION OF THE AUDIBLE, VISIBLE AND BIMODAL ATTRIBUTES OF TALKING AND SINGING FACES

David J. Lewkowicz

NYS Institute for Basic Research
1050 Forest Hill Rd.
Staten Island, NY 10314

ddlsi@cunyvm.cuny.edu

ABSTRACT

Human faces and accompanying voices are a ubiquitous part of the infant's perceptual experience. They serve as a major vehicle for the acquisition of linguistic, social, and emotional skills, just to name a few. Therefore, the study of how infants perceive the audible, visible and combined attributes of face/voice compounds can show how various features of face/voice compounds contribute to the acquisition of higher-level skills. I review the findings from a series of studies in which we investigated infants' response to a variety of features of face/voice compounds. These findings show that there are important developmental differences in the way infants respond to the audible, visible and bimodal components of face/voice compounds and that these differences are dependent on the specific nature of the information given. Overall, the results suggest that infants, like adults, may respond to multiple sources of information when attempting to evaluate the "message" carried by bimodal speech but that their functional use of this information is limited by the infant's developmental status.

1. ISSUES AND QUESTIONS

Talking or singing faces are relatively ubiquitous in infants' daily lives. From a perceptual point of view, they provide a set of complex and hierarchically organized features that must somehow be perceived to get to the "message." One way to get to the message is to integrate the diverse auditory and visual sources of information into unified objects/events. By integrating these heteromodal sources of information, the perceiver no longer merely perceives a collection of diverse, heteromodal features. As a result, perception is more efficient, easier, and more adaptive. In general, intermodal integration is a process that makes it possible for the perceptual system to relate two or more heteromodal inputs. Integration can include one of three distinct kinds of processes: (a) non-specific effects of stimulation in one modality on responsiveness to stimulation in a different modality, (b) matching of equivalent heteromodal features such as intensity or overall temporal rhythm (also known as amodal perception), or (c) association of modality-specific attributes (e.g., the timbre of a person's voice and the configuration of their face). It should be noted that this definition of integration differs from the way some writers define intermodal integration, particularly when referring to bimodal speech perception. For example, for Burnham [1] integration refers to the fusion of the audible and visible attributes of linguistic information, often resulting in a new percept (see discussion of the McGurk effect in Section 2). Because the principal aim of the current article is to place the work on infants' response to talking and singing faces in a broader intermodal perception framework, the definition of integration offered here will be used.

In order to integrate the cues inherent in talking and singing faces, infants must rely on their sensory, perceptual, and cognitive skills. These skills are relatively limited in early human development by the fact that the auditory and visual modalities are immature at birth and that sensory, perceptual, and cognitive processes undergo significant developmental changes and re-organization during infancy [2-4]. Complicating the picture further is the fact that the structural and functional development of the different sensory systems does not occur all at the same time. Although the development of all the sensory systems begins during the prenatal period, the somesthetic/vestibular modality is the first to develop, followed in order by the chemical, the auditory, and the visual modalities [5]. This sequential functional onset of the sensory systems means that at birth each sensory system starts out life with a different experiential history and this is likely to have a major impact on subsequent postnatal perceptual development and differentiation. This, in turn, may mean that integration of heteromodal sources of input may be more or less difficult at a given point in development depending both on the type of experience that the specific modality might have had up to that point and on the extent of its structural and functional maturity.

2. THEORETICAL VIEWS ON THE INTEGRATION OF HETEROMODAL SOURCES OF INFORMATION

As indicated earlier, perception of objects and events that are concurrently specified by heteromodal sources of information is more efficient and adaptive when they can be perceived in an integrated manner. Indeed, evidence suggests that adult subjects do not perceive talking faces in terms of their separate audible and visible attributes but as integrated wholes [6]. This raises an interesting and important question from the standpoint of development. Given that the different sensory systems are immature during early development and that they have limited processing capacity, how might infants perceive multimodally specified objects and events? The initial answer to this question came from theoreticians. One of the earliest modern theoretical statements specifically concerned with development and attempting to answer this question came from Piaget [7]. His view of development was constructivist in nature and thus assumed that various sensory-motor, perceptual, and cognitive functions are slowly elaborated and constructed through the infant's experience and interaction with the world. As a result, Piaget assumed that the senses operate as separate avenues of sensory input at birth and that slowly, over many months or even years, begin to communicate and cooperate with one another. In contrast, J. J. and E. Gibson [8-10] held that perception does not consist of separate sensations in the different sensory modalities. Instead, the Gibsons argued that humans perceive objects and events as integrated wholes and that this kind of integrated perception is based on the ability to perceive amodal invariants (i.e., properties that can be specified equivalently in the different sensory modalities). With specific regard to early development, E. Gibson proposed that infants are sensitive to amodal invariants right from birth and that as they develop they become increasingly more sensitive to increasingly finer amodal invariant structure. This would suggest that when infants are confronted with a talking face, they attend primarily to the invariant amodal attributes of that talking face. For example, rather than focusing attention on the color of a speaker's eyes or hair and/or on the pitch of the speaker's voice, E. Gibson's view would suggest that infants would primarily focus on the speaker's overall tempo, rhythm, affective tone, and spatial co-incidence. From the standpoint of bimodal speech perception these might be some of the properties that can specify a given speech utterance equally well in the auditory and visual modalities.

The empirical evidence that has been amassed since Piaget's and the Gibsons' influential views were

proposed actually provides some support for each view. On the one hand, there is a good deal of evidence to support the claim that infants respond to amodal invariants. For example, Lewkowicz and Turkewitz [11] showed that neonates can equate auditory and visual inputs on the basis of what is probably the most basic of amodal invariants: intensity. It has also been shown that infants can perceive intermodal relations based on synchrony [12-16] and duration [12], and that they can perceive the intermodal relation between the auditory and visual aspects of affect [17] and of syllables [18]. On the other hand, evidence indicating that infants respond to amodal invariants does not mean that responsiveness to intermodal relations based on equivalence is the only and primary way in which infants perceive their world when faced with multimodally represented objects or events. First, there is evidence that infants find it difficult to respond to certain amodal invariant relations. For example, infants find it difficult to perceive intermodal equivalence based on temporal rate [19-22]. Second, as will be shown here, the evidence from our studies on infants' response to bimodal speech is not fully consistent with the amodal invariance detection view. Our work makes it evident that part of the problem lies in focusing too much attention on the detection of amodal invariants to the exclusion of the other two processes that were included earlier in the definition of intermodal integration. The ultimate aim in science is to try to find an organizational scheme that identifies some common and more basic process that underlies what might appear to be qualitatively different processes. Is it possible to identify some common process that can provide a conceptual "glue" for what appear to be qualitatively different intermodal perceptual processes?

One candidate process that has recently been offered that might provide this conceptual "glue" is perceptual recognition. based on multiple sources of information. Massaro [6] has argued that perceptual recognition based on multiple sources of information is part of a general process of pattern recognition. According to Massaro, perceptual recognition is dependent on the processes of evaluation, integration, and decision and each of these, in turn, is dependent on prototypes stored in long-term memory. When different sources of information (e.g., various cues associated with the audible and visible attributes of speech) are initially evaluated they are transformed into psychological values and these values are then integrated and used to decide between various speech alternatives. The response decision that enables a perceiver to decide between different likelihoods that a given alternative (e.g., a /da/ or a /ba/) is likely is based on the output of the integration process. Massaro's model makes several key assumptions. First, each source of

information is evaluated and a determination is made about the continuous degree to which that source corresponds to different alternatives. Second, the sources of information are assessed independently of one another. Third, the sources are integrated and this provides a continuous degree of support for each alternative. Fourth, the identification of a given percept is based on the relative amount of support among the alternatives.

From a theoretical perspective, Massaro's model certainly might be able to account for the kinds of intermodal perceptual abilities observed in infants. The problem is, however, that the nature of the data from the studies that are currently available does not permit a direct evaluation of the utility of the model for developmental phenomena and of its ability to account for the diverse findings in the developmental literature. The findings from our studies on infants' response to talking and singing faces also cannot be used to directly test the utility of Massaro's model because their design is also not appropriate for testing the model. As a result, the findings to be reviewed will be used to determine which of the theoretical accounts discussed so far might be able to offer the most useful general principles that might account for the effects found. Before doing so, however, it might be useful to consider what may be involved in the integration of the audible and visible information specifying a speech utterance.

The perception of talking faces is part of a more general process of intersensory integration. If one adopts Massaro's approach to the perception of multimodal sources of information, the perceiver must be able to integrate all the different cues in order to correctly identify the "message" inherent in a talking face. One way to accomplish the integration is to do it on the basis of some of the most basic amodal invariants inherent in a talking or singing face: space and time. That is, the audible and visible attributes of talking and singing faces correspond simply because they occur in the same space and time. Indeed, research on intersensory integration indicates that adults rely on spatial and temporal coincidence to unify their multimodal world [23]. When that coincidence is disrupted, adaptation and recalibration effects are observed and the perceptual system attempts to make discordant inputs concordant.

Two widely known examples of the powerful proclivity for unified intermodal perception and for the importance of spatiotemporal factors for intermodal integration are the McGurk and the ventriloquism effects. In the McGurk effect [24], a subject's auditory perception of a syllable is altered by a conflicting visual syllable presented at the same space and time. In the ventriloquism effect, the location of a sound that is spatially separate from a visual stimulus is actually perceived to be closer to the visual stimulus than it really is [25-27]. With specific regard to talking faces, Massaro [6] has shown that adult subjects normally integrate the audible and visible features of talking faces to make a perceptual judgment. Our work on infants' response to talking faces also suggests that infants might engage in a similar process but with a twist. Probably because of a combination of sensory/perceptual/cognitive limitations, and the relative lack of perceptual experience, infants sometimes do not respond to all available inputs and thus do not appear to integrate audible and visible inputs the same way that adults do. These and other types of findings are described in detail in the next section.

3. INFANTS' PERCEPTION OF TALKING FACES

There is no doubt that infants can integrate the information specified by the face and voice. For example, it has been shown that infants as young as 3-4 months of age can perceive the equivalence between a face and a voice producing isolated audio-visual syllables [18], entire passages [28], or expressions of affect [17, 29]. It even appears that infants' perception of the audible aspects of speech is influenced by its visible aspects in a manner consistent with the McGurk effect [30], although Burnham [1] argues that these results do not support the conclusion that infants fuse the audible and visible information to produce a novel percept. He does, however, report results that suggest that infants do perceive the McGurk effect in a fused manner.

Given the earlier-discussed fact that intermodal integration includes a diverse set of processes, a full understanding of the development of intermodal perceptual abilities should include investigations whose design allows for the evaluation of other effects besides the detection of amodal invariants. In other words, it would be useful to conduct studies whose design makes it possible to evaluate the contribution of information in each modality to the perceptual response. The studies reviewed here do have this design feature. They represent a systematic program of research that investigates infants' response to face/voice compounds that is motivated by the desire to better understand the contribution of unimodal as well as bimodal information to perception. One unique aspect of the design of these studies is that it not only permits the evaluation of the separate contribution of unimodal and bimodal inputs to responsiveness but also whether infants can detect the intermodal relations inherent in the face/voice compound stimulus. The results from these studies provide an empirical description of the

ways in which infants respond to the components of face/voice compounds. In addition, they provide an opportunity to determine how well the different theoretical approaches to the development of intermodal perception outlined earlier can account for the results. On the one hand, the amodal invariance detection view would suggest that infants should respond to face/voice compounds as unified objects/events. On the other hand, Massaro's model, when applied to developmental phenomena, might be interpreted to mean that organisms who are not fully mature and whose different sensory systems are not functionally equivalent might respond only to as many sources of information that their sensory systems might permit. The developmental version of this latter view would admit to the possibility that non-integral types of responsiveness might be possible as well.

Based on the assumption that perception is based on the evaluation of all sources of information, Massaro [6] has argued that in order to understand how facial and vocal information contributes to the perception of talking faces one must assess separately the subject's response to the auditory, visual, and bimodal information. Our experimental approach is based on a similar view. To do so, we have used what I have dubbed the multimodal component variation method. Although this method is similar to the type of experimental method favored by Massaro, it differs from it in some important respects. Whereas Massaro and his colleagues typically give dozens of trials to each subject, we cannot do so because this is impossible with infants. In addition, in order to answer questions regarding sensory/perceptual responsiveness in infants, one must use methods that capitalize on their limited response capacities. The habituation/test procedure, which has been used widely in studies of infant perception, is one such experimental procedure.

3.1. The Habituation/Test Procedure

The habituation/test procedure involves presenting a given stimulus (e.g., a talking face) repeatedly over several trials. This can be done either by presenting the stimulus for a pre-determined number of trials or until an infant reaches a pre-defined habituation criterion. In either case, the aim is to enable the infant to become familiar with the habituation stimulus. Once the habituation phase ends, a set of test trials is given during which novel information is presented. The object of this procedure is to determine whether the infant can discriminate the difference between the habituation stimulus and the novel stimulus presented during the test phase. Typically, response recovery is interpreted to mean that the infant discriminated the difference between the familiar and novel stimuli and thus detected the novel information.

3.2. The Multimodal Component Variation Method

The multimodal component variation method is simply a particular way of implementing the habituation/test procedure in order to examine responsiveness to audio/visual compound stimuli. Thus, following the habituation phase when infants see and hear the same talking face over a number of trials, they are then given four types of test trials: auditory (A), visual (V), auditory/visual (A/V) and familiar (F). During the A test trial some aspect of the auditory component is changed while the visual component remains unchanged. During the V test trial some aspect of the visual component is changed while the auditory component remains unchanged. During the A/V test trial some aspect of both components is changed. Finally, during the F test trial the habituation stimulus is presented again and the response to this stimulus is used as the baseline against which response recovery in each of the other three test trials is assessed. If the duration of looking in any of the other three test trials is significantly greater than the duration of looking in the F test trial, then this is taken to mean that the infants detected and discriminated the change on that particular test trial.

Although not central to the experiment itself, two other types of trials are given: a single pre-test and a single post-test trial during which infants can see and hear a segment of a Winnie-the-Pooh cartoon. The pre-test trial is given as the very first trial and the post-test trial is given as the very last trial. These two trials are critical for determining whether fatigue might have contributed to the failure to respond in the test trials. That is, if infants do not exhibit response recovery in the A, V, or A/V test trials but do exhibit response recovery in the post-test trial and that response does not differ from the response in the pre-test trial then failure to respond in the test trials cannot be ascribed to fatigue. Put differently, this kind of result suggests that infants failed to discriminate the specific change presented in that specific test trial.

It should be noted that the multimodal component variation method differs from Massaro's method in that responsiveness to a unimodal change is studied *in the presence* of input in the other modality. This is based on the implicit assumption that in order to understand how infants respond to the components of bimodal compounds, one should study responsiveness in a bimodal context. Moreover, given the use of the habituation/test procedure which is explicitly designed to determine whether an infant can discriminate between two events, one cannot habituate an infant to a bimodal compound stimulus and then present a changed component in the absence of the other component. Doing so poses

the interpretive problem of not being able to determine whether a successful discrimination in such a case is due to the detection of the component change *per se* or to the detection of the change from bimodal to unimodal stimulation.

One of our concerns when we began this work was that the information inherent in the face of a person engaged in the act of speaking a continuous utterance was so complex that infants might have a hard time detecting changes in a single dimension. This kind of finding would be interesting in and of itself but would be difficult to interpret given that it would be a negative finding. As a result, our initial strategy was to insure that we obtained some positive findings first and then go on and systematically manipulate the various cues to determine which ones actually controlled responding. Given this strategy, we decided to study responsiveness to concurrent variations in several dimensions.

3.3. Perception of the Audible, Visible and Bimodal Attributes of Bimodal Speech

In most cases when infants interact with another social partner they are exposed to that partner's continuous bimodal speech. This fact raises two interesting questions. First, to which aspects of such speech might infants be responsive? Second, given the dramatic changes in sensory/perceptual abilities observed during the first months of life [2, 3], might there be developmental changes in the way infants respond to bimodal speech and its constituents? The first study was part of a series of studies [31] and was aimed at answering these questions. Infants were habituated to a talking face reciting a continuous script and then tested to audible, visible, and bimodal changes to determine what aspects of this continuous, bimodal speech infants perceived.

Separate groups of 4-, 6-, and 8-month-old infants, composed of 32 infants each, were the subjects for this study. These ages were chosen to span a period in early development where, as indicated earlier, a variety of intermodal perceptual abilities begin to emerge because it is during this time that responsiveness to the face as a social object emerges and because speech perception undergoes important changes. Each infant was given a total of 18 trials consisting of a pre-test trial, 12 habituation trials, the 4 types of test trials described earlier (presented in counterbalanced order across infants at each age), and a post-test trial. During the habituation phase each infant could see and hear a video segment showing a female reciting a continuous utterance for 20 s every 2-3 s. While reciting the utterance the female spoke in an adult-directed (AD) manner, meaning that she spoke in a manner that most adults use to speak to one another (i.e., relatively

monotone and unanimated). As indicated earlier, in order to maximize the possibility of discrimination, infants were given changes along several dimensions simultaneously. As a result, during the A test trial a novel female could be heard reciting a novel continuous utterance and the familiar female could be seen mouthing the familiar utterance. During the V test trial a novel female could be seen mouthing a novel utterance and the familiar female could be heard uttering the familiar utterance. Finally, during the A/V test trial a novel female could be heard and seen reciting a novel utterance. While the infant watched and listened to the talking face, we videotaped his or her visual fixations and scored the duration of looking at the video monitor during each trial.

Infants at each age exhibited reliable habituation. In addition, as can be seen in Figure 1, infants of all three ages exhibited significant response recovery in the post-test trial indicating that they were not fatigued (significant response recovery in the post-test trial was found in all of the subsequent studies to be discussed here and, thus, this result will not be discussed again and only the data for this trial will be presented in each subsequent figure). An overall analysis of variance was first performed on all the test trial data and then the analysis was followed up with a series of planned-comparison tests,
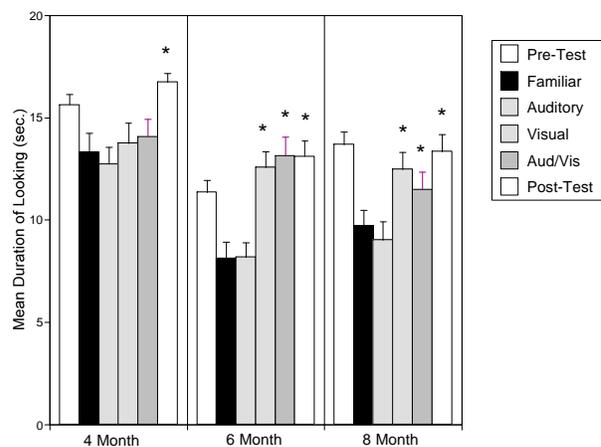

Figure 1. Mean duration of looking at auditory, visual, and bimodal component changes in 4-, 6-, and 8-month-old infants habituated to a female speaker uttering a script in the AD manner and tested with components of a different female speaker also speaking in the AD manner. Asterisks indicate significant response recovery to the type of change indicated by the bar underneath.

comparing the duration of looking in the F test trial versus the duration of looking in the A, V, and A/V test trials, respectively. The planned comparison tests were justified by prior work in our laboratory with auditory/visual compound stimuli composed of a spatially dynamic visual stimulus and a concurrent

sound that was related in various temporal ways to the visual stimulus [13]. In these studies we found that infants as young as two months of age had no difficulty in discriminating the two unimodal as well as bimodal changes in the compound stimulus. As a result, we expected that infants in the current study should have no problem responding to the unimodal and bimodal changes.

As can be seen in Figure 1, the 4-month-old infants did not discriminate any of the component changes despite the fact that they did exhibit significant response recovery in the post-test trial. In contrast, both the 6- and 8-month-old infants discriminated the visual and the combined component changes. These results suggest that infants younger than 6 months of age have difficulty perceiving the audible and visible information inherent in a talking face when the information in each modality represents a continuous utterance spoken in the AD manner.

In an effort to further expand our exploration of the features that infants might be able to perceive, we turned our attention to gender. This is one of the most ubiquitous cues that differentiates between talking faces and it is one of the cues that infants are exposed to right from birth. Moreover, studies have shown that infants are sensitive to gender differences [32]. Therefore, it might be reasonable to expect that they might be sensitive to gender differences inherent in talking faces. To determine if that is the case, we repeated the first study by using the same methods and procedures and tested new groups of 4-, 6- and 8-month-old infants (32 infants/group) [31]. In this study, infants were habituated to a male talker speaking the same utterance as in the first study and tested with component changes produced by a female talker speaking the other utterance. Both speakers recited the utterance in the AD manner.

As can be seen in Figure 2, a gender difference made it possible for the youngest infants to discriminate the bimodal and visible changes. As in the first study, the two older groups of infants once again discriminated the visible and bimodal component changes. These results indicate that gender is a salient visible characteristic of bimodally specified talking faces to infants as young as four months of age.

One of the interesting and to some extent unexpected findings from the first two studies was that infants of all three ages did not discriminate the audible component changes. What makes this finding so interesting is that infants did not respond to the vocal changes despite the fact that the lexical/syntactic content of the script changed, the identity and gender of the person changed, and the vocal and audible components of the face became desynchronized in that test trial. This outcome may
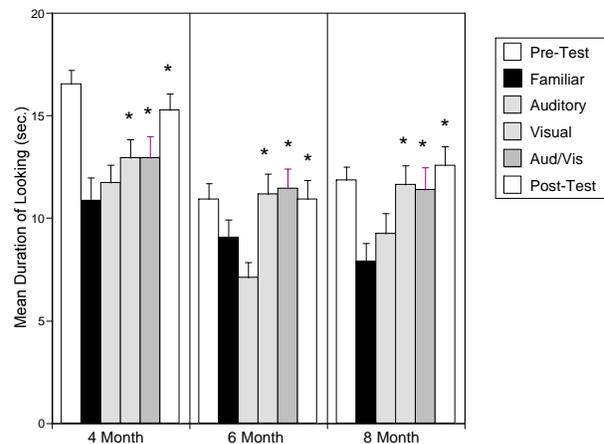

Figure 2. Mean duration of looking at auditory, visual, and combined component changes in 4, 6, and 8-month-old infants habituated to a male speaker and tested with components of a female speaker. Asterisks indicate significant response recovery to the type of change indicated by the bar underneath.

be due either to an insensitivity of the experimental procedure to detect infants' true auditory discriminative ability or to the fact that the auditory change was not salient enough. To investigate these possibilities, the salience of the vocal component was increased further in the next study by introducing infant-directed (ID) speech. This type of speech is what adults often use when speaking to infants. Its most salient characteristics are highly modulated frequency contours, a slower overall tempo, and exaggerated temporal structure consisting of elongated pauses. Research has shown that infants prefer ID speech over AD speech [33-36].

To determine whether infants might be able to detect features of the audible component when the differences in that component are along the manner-of-speech dimension, we used the identical test procedures as in the first study. This time, however, we habituated separate groups of 32 infants aged 4, 6, and 8 months to a male speaker uttering a script in the AD manner and then tested them with aspects of a different script spoken by a female who spoke in the ID manner [31].

As can be seen in Figure 3, introducing a manner-of-speech difference on top of a gender difference made it possible for the first time to obtain significant discrimination of the audible change in the two older groups of infants. Thus, both the 6- and 8-month-old infants not only responded to the bimodal and visible component changes, but they responded to the audible component change as well. In contrast, the 4-month-old infants still only responded to the bimodal and visible component changes. These findings show that both 6- and 8-

month-old infants are sensitive to audible changes in the manner of speech in a bimodal context when those changes involve a change from AD to ID speech together with gender changes. These findings also show that the failure to exhibit significant response recovery in the A test trial in the previous studies was not due to the insensitivity of the visual attention measure to changes in auditory information.
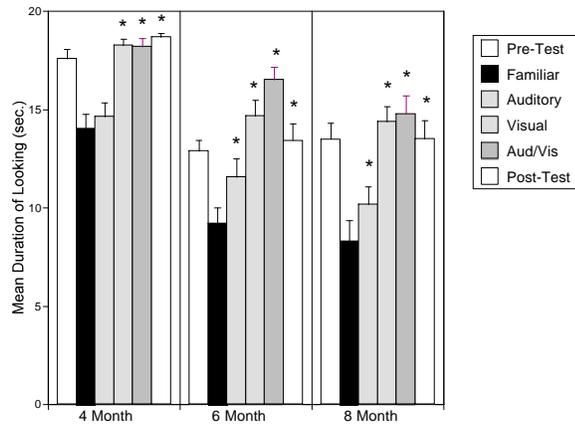

Figure 3. Mean duration of looking at auditory, visual, and combined component changes in 4, 6, and 8-month-old infants habituated to a male speaker speaking in the AD manner and tested with components of a female speaker speaking in the ID manner. Asterisks indicate significant response recovery to the type of change indicated by the bar underneath.

It was pointed out earlier that an underlying assumption of the component variation method is that in order to understand infants' perception of bimodal events, responsiveness to each component making up the bimodal event must be studied in the context of the other component. Consistent with this view are the results of several studies on infants' responsiveness to differences between ID and AD speech. For example, studies have found that infants exhibit a significant preference for ID speech right from birth when the ID speech is presented against a constant, non-face-like visual background such as a checkerboard or when it is presented with no specific visual stimulus [33, 34]. Likewise, other studies have found that 7-week-old infants can discriminate ID from AD speech when the speech segments are presented against the background of a constant checkerboard pattern but that they do not prefer ID speech over AD speech when it is bimodally specified [37]. In contrast, infants aged 4-5.5 and 7.5-9 months of age prefer to look at a person speaking in the ID manner over a person

speaking in the AD manner as long as the ID speech is bimodally specified. In other words, infants' responsiveness to ID speech differs depending on whether the visual information that accompanies the ID speech is a face or not and on the infant's age. Early in development infants appear to prefer ID speech when it is specified only in the auditory modality whereas later in development they exhibit the preference only when it is specified bimodally. These kinds of findings suggest that the information conveyed by the face, and/or the relation that the auditory information bears to the facial information plays an important role in responsiveness and should not be ignored.

Some of the developmental differences in responsiveness to unimodally vs. bimodally represented ID speech probably reflect the differential onset of function in the auditory and visual modalities and thus of their differential functional differentiation. As noted earlier, the auditory modality begins to register input earlier (during the last trimester of pregnancy) than does the visual modality (after birth) and evidence indicates that auditory experience during the last trimester affects the way newborn infants respond to auditory stimulation [38]. Thus, the developmentally earlier auditory preference for ID speech may be related to the fact that the auditory modality is more experienced. In the same vein, the later preference for bimodal ID speech may reflect the later and gradual maturation of visual function, the gradual acquisition of visual experience, and the gradual integration of visual and auditory processing. These kinds of considerations make it clear that it is important to study responsiveness to the unimodal components making up a bimodal compound stimulus in the context of the other component.

Table 1 summarizes the results from the first three studies and shows the developmental differences in responsiveness to the various stimulus dimensions. When infants were habituated to bimodal speech spoken in the AD manner and then tested with changes also spoken in the AD manner (spoken by a different person reciting a different utterance), only the two older groups responded and only to the V and to the A/V changes. When a gender difference was added, 4-month-old infants also detected the V and A/V changes. When a manner-of-speech difference was combined with a gender difference the two older groups of infants now also detected the A component change.

3.4. The Role of Exaggerated Prosody

	A	V	A/V
Type of Change			
AD→AD			
4 month	-	-	-
6 month	-	X	X
8 month	-	X	X
AD→AD & Gender			
4 month	-	X	X
6 month	-	X	X
8 month	-	X	X
AD→ID & Gender			
4 month	-	X	X
6 month	X	X	X
8 month	X	X	X

Table 1. A comparison of the outcomes of the first three studies investigating the discriminability of the unimodal and bimodal features of continuous utterances and how it is affected by gender and manner of speech differences. A **-** means a failure to discriminate that particular component change whereas an X

As expected, the manner-of-speech feature turned out to be important for the perception of talking faces. Another highly salient and universal communicative tool that adults use with infants is singing [39]. One of the reasons that adults sing to infants is because infants are highly responsive to singing. One of the unique and singular features of singing is that it, even more so than ID speech, is characterized by exaggerated prosody variations. If infants are particularly tuned to exaggerated prosody variations then it might be reasonable to expect that they would be highly responsive to singing. Moreover, perhaps responsiveness to singing might emerge earlier in development than to the exaggerated prosody typically found in ID speech.

To test this possibility, we conducted another study [40] in which we utilized the same methods as before. To find out whether aspects of singing might lead to the earlier developmental emergence of responsiveness to attributes of face/voice compounds as well, we added a 3-month age group to the study. Thus, we tested separate groups of 32 infants each, aged 3, 4, 6, and 8 months. During the

habituation phase we presented a female speaking the same utterance (in the AD manner) as in the previous studies. During the test phase we tested infants with the attributes of a different female who sang a portion of a song from the Broadway show "Oklahoma."

As can be seen in Figure 4, all four age groups responded to the bimodal change and to the visual change. In contrast, and as predicted, not only did the 6- and 8-month-old infants respond to the audible change, but now even the 4-month-old infants responded to it. Thus, it appears that the greater prosody cues inherent in a song were salient enough to be perceived even earlier in development. Interestingly enough, however, the prosodic cues inherent in singing did not appear to be sufficiently salient to the 3-month-old infants as they did not exhibit significant response recovery.
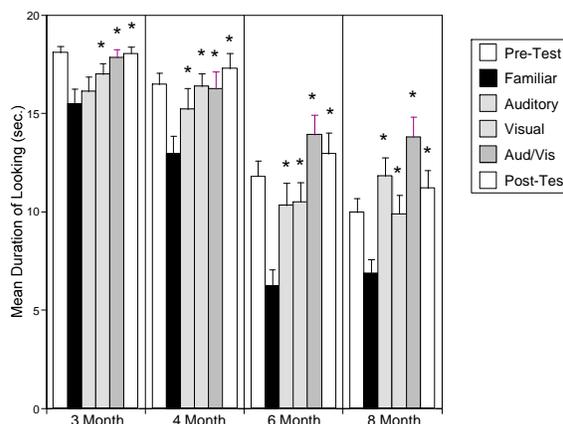

Figure 4. Mean duration of looking at auditory, visual, and combined component changes in 3-, 4-, 6-, and 8-month-old infants habituated to a female who spoke in the AD manner and tested with components of another female who sang. Asterisks indicate significant response recovery to the type of change indicated by the bar underneath.

Given that the previous studies showed that gender is a discriminable feature, we repeated the speech-singing study except that this time we added a gender difference [40]. Again, separate groups of 32 infants each, aged 3, 4, 6, and 8 months were habituated to a male uttering the same script as before in the AD manner and then were tested with attributes of a female who sang. As can be seen in Figure 5, this time all age groups discriminated all three types of changes. In other words, the addition of a gender difference made it possible even for the 3-month-old infants to respond to the audible change. This developmental pattern is similar to the pattern obtained earlier where the addition of a gender difference enabled the 4-month-old infants to respond to bimodal and visible component changes

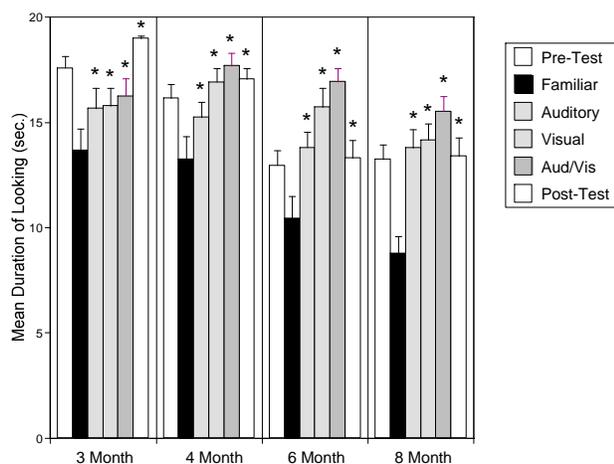that they did not respond to in the absence of the gender difference.

Figure 5. Mean duration of looking at auditory, visual, and combined component changes in 3-, 4-, 6-, and 8-month-old infants habituated to a male who spoke in the AD manner and tested with components of a female who sang. Asterisks indicate significant response recovery to the type of change indicated by the bar underneath.

Table 2 shows a summary of the results of the five studies discussed so far. It shows clearly that as more features are added, the developmental emergence of responsiveness occurs earlier and to more component changes. Moreover, as Table 2 makes clear, one of the most interesting and consistent findings from all the studies reviewed so far is the fact that the most difficult discrimination for infants to make was the audible one. Although by itself this pattern of findings may not appear to be surprising, it is somewhat curious when considered in light of findings from other studies on infants' responsiveness to auditory/visual compound stimuli. For example, in other studies utilizing the multimodal component variation method, I have found that infants have little difficulty discriminating the changes in the audible component. This was the case both in studies in which infants were tested with changes in the components of a compound stimulus consisting of a flashing checkerboard and a pulsing tone [41, 42] as well as with changes in the components of a compound stimulus consisting of a bouncing disk and a synchronous tone [13]. These findings suggest that the poorer discrimination of the audible changes in the bimodal speech studies might be, in part, due to the specific nature of the stimulation presented. In particular, inattention to the auditory information might have been the result of either the complexity of the speech utterance, the fact that faces might be extremely captivating, or a combination of the two .

Type of Change	A	V	A/V
AD→AD			
4 month	-	-	-
6 month	-	X	X
8 month	-	X	X
AD→AD & Gender			
4 month	-	X	X
6 month	-	X	X
8 month	-	X	X
AD→ID & Gender			
4 month	-	X	X
6 month	X	X	X
8 month	X	X	X
AD→Song			
3 month	-	X	X
4 month	X	X	X
6 month	X	X	X
8 month	X	X	X
AD→Song & Gender			
3 month	X	X	X
4 month	X	X	X
6 month	X	X	X
8 month	X	X	X

Table 2. Outcomes of studies in which responsiveness to differences between the unimodal and bimodal features of continuous utterances was investigated. A - means a failure to discriminate whereas an X means a successful discrimination.

Continuous utterances such as those used in the studies so far discussed contain many different cues at hierarchically distinct levels. Among these are phonemic, syllabic, lexical, and syntactic ones, to name a few. Each is a potentially useful cue for identification of the spoken message but may not

necessarily be used, particularly when the receiver's ability to process the information is not mature. In other words, although infants' speech processing skills are quite impressive and are constantly improving during the first year of life [43], their linguistic skills lag far behind and take many years to develop. Thus, it is not all that surprising that the infants in the studies so far described did not respond to the audible changes. Although infants might be able to discriminate audible changes on the basis of acoustic variations, it may be that the various higher-level cues made the detection of acoustic or other lower-level features difficult due to masking or some other such process. One way to put this possibility to test is to simplify the information by presenting single, isolated syllables.

3.5. Perception of Bimodal Syllables

To study responsiveness to single syllables we used the multimodal component variation method once again. This time, however, we used an infant-controlled habituation/test procedure [44]. The principal feature of this method is that it tailors the number of habituation trials to each infant's individual processing style. To implement this, each infant was permitted to control the duration of stimulus presentation on each trial (up to a maximum of 32 sec.). The infant's looking was observed and each time the infant looked at the video monitor, he could see and hear the person uttering the syllable; each time he looked away from the monitor for more than one second, the presentation ceased and the trial ended. Habituation was deemed complete when the infant's looking duration during the last three (i.e., terminal) trials decreased by 50% relative to the total amount of looking during the initial three trials. This criterion was calculated by way of a moving window such that the terminal set of 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 trial. This sliding window criterion continued to be applied until the 50% criterion was met. The test phase commenced once this criterion was met. During the test phase the duration of each trial also was controlled by the infant's looking behavior, with the maximum duration set to 32 s.

Groups of 32 infants aged 4, 6, and 8 months were habituated to a videotaped segment of one of two females speaking one of two syllables (\ba\ or \sha\) in the AD manner. They were then tested with the same female uttering the other syllable also in the AD manner. The two syllables were chosen to maximize the contrast between the familiar and novel syllables and thus were clearly distinct in terms of both their audible and visible properties

(voicing, place of articulation, and manner of articulation). The reason that the same female was used during both phases of this study was to minimize the number of discriminative cues and thus begin to zero in on the specific cues underlying discrimination. Consequently, during the A test trial infants saw the familiar female mouthing the familiar syllable and heard the same female uttering a novel syllable. In the V test trial they heard the familiar female uttering the familiar syllable and saw the same female mouthing a novel syllable. Finally, during the A/V test trial infants saw and heard the familiar female uttering the novel syllable.

Figure 6 shows the results from this study. As can be seen, all three age groups discriminated the combined change and the audible change. Unexpectedly, however, neither the 4-month-old nor the 6-month-old infants discriminated the visible change. This is in marked contrast to the results from the continuous utterance studies where responsiveness to the visible changes was relatively robust. One possible reason that might account for the failure to discriminate the visible change may be the fact that the person's face did not change; only the motion of the lips changed. Moreover, whereas in the current study the person in the habituation and the test phases was seen mouthing two different syllables, in the continuous utterance studies the two people seen in the two phases were seen mouthing two different continuous passages. It may be that the visible differences specifying two syllables spoken by the same person are more difficult for infants to discriminate than the visible differences specifying different samples of continuous speech spoken by different people.
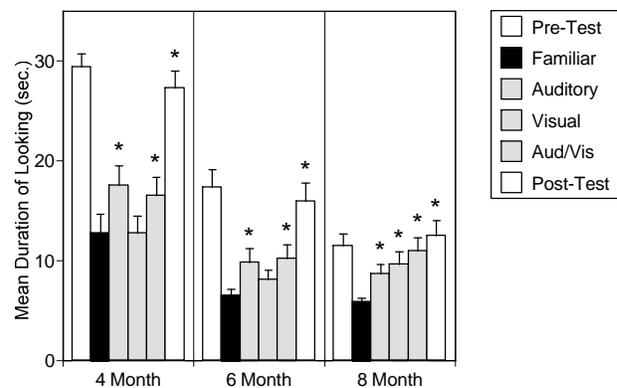

Figure 6. Mean duration of looking at auditory, visual, and combined component changes in 4, 6, and 8-month-old infants habituated to a female speaker uttering a single syllable in the AD manner and tested with components of the same female uttering a different syllable. Asterisks indicate significant response recovery to the type of change indicated by the bar underneath.

If the failure of the 4- and 6-month-old infants to detect and discriminate the visible change was due to the relative difficulty of detecting the visible contrast between the two syllables, one way to enhance it is to add exaggerated prosody. We did this in the next study [44]. We habituated groups of 32 infants aged 4, 6, and 8 months to a female speaking one of the two syllables in the AD manner and then tested them to changes to the other syllable spoken by the same person except that this time she spoke the syllable in the ID manner. In this case, when she spoke she appeared to be visibly more animated and could be seen pronouncing each instance of the syllable in an exaggerated and slower manner. We hypothesized that these more exaggerated visible features would be more easily discriminable and that the 4- and 6-month-old infants would be able to perform the discrimination.

Figure 7 shows the results from this study and indicates that the exaggerated visible features did, indeed, facilitate discrimination. In contrast to the previous study, the infants in this study discriminated not only the bimodal and audible
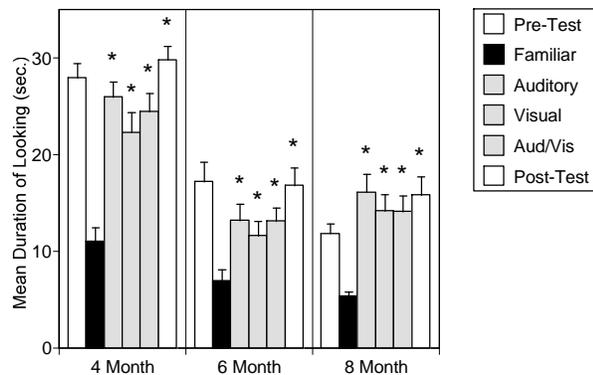
Figure 7. Mean duration of looking at auditory, visual, and combined component changes in 4, 6, and 8-month-old infants habituated to a female speaker uttering a single syllable in the AD manner and tested with a different syllable spoken in the ID manner by the same female. Asterisks indicate significant response recovery to the type of change indicated by the bar underneath.

changes, but the visible ones as well. Table 3 is a summary of the results from the first two syllable studies. It shows that when infants were habituated to a single syllable spoken in the AD manner and then tested with changes also spoken in the AD manner (spoken by the same person reciting a different syllable), the two younger groups responded only to the A and to the A/V changes, whereas the 8-month-olds responded to all three types of changes. When a manner-of-speech difference was introduced, the two youngest groups of infants also discriminated the V change.

Type of Change	A	V	A/V
AD→AD			
4 month	X	-	X
6 month	X	-	X
8 month	X	X	X
AD→ID			
4 month	X	X	X
6 month	X	X	X
8 month	X	X	X

Table 3. Summary of the syllable studies in which responsiveness to differences between the unimodal and bimodal features of the syllables \ba\ or \sha\ was investigated. A - means a failure to discriminate whereas an X means a successful discrimination.

The factor that most likely made it possible for the infants in the AD→ID experiment to discriminate between the visible attributes of the face seen during the habituation phase and the face seen during the test phase was motion. In general, motion plays an important role in infants' response to auditory/visual compounds. For example, when infants' responsiveness to the auditory and visual components of a compound stimulus is tested, responsiveness to the visual component is affected by whether this component is spatially static or dynamic [41, 42]. Thus, when infants are tested with a compound stimulus composed of a spatially static flashing checkerboard and a pulsing tone, infants as old as 6 months of age do not respond to changes in the flash rate of the checkerboard. In contrast, when tested with a compound stimulus composed of a moving and sounding stimulus, infants as young as 2 months of age respond to changes in the bounce rate of the visual component. In addition to these findings, the findings from the speech studies described above show that when motion is available infants have little difficulty in discriminating the visible changes.

Of course there are different types of motion and they affect responsiveness in different ways [45, 46]. For example, in terms of the overall degree of motion, a singer projects a good deal of motion and more than a person engaged in the act of speaking. When speaking, a person's face, lips and head can produce different degrees of visual motion

depending on the specific type of speech that the person produces. Perhaps the least kinematically 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.

The absence of clear kinematic differences might have been one reason why infants had difficulty making the visible discrimination in the first syllable study. A second reason may have been the 4- and 6-month-old infants' insensitivity to differences in the shaping of the lips during the production of the syllables. This conclusion is supported by findings that object perception undergoes a major shift around the seventh month of life [47]. Before seven months of age object perception is based on an edge-insensitive process. In other words, before seven months of age infants are not sensitive to edge relationships that normally specify objects and are instead responsive primarily to their kinematic properties. At around seven months of age, however, the edge-insensitive process is replaced by an edge-sensitive process. This 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. As a result, by seven months of age infants become responsive both to the kinematic properties of objects and to their edge information and thus can perceive objects in a more detailed and accurate fashion. This kind of perceptual shift, together with the general process of perceptual differentiation [9, 48] most likely accounts for the age difference in infants' response to the visible syllable changes found in the first syllable study. Specifically, the kinematic differences between the two syllables were probably not sufficiently distinct for the 4- and 6-month-old infants in the first syllable study. Moreover, these infants most likely did not perceive the finer visual aspects of the two syllables because the edge-sensitive mechanism had not developed yet. In contrast, the 8-month-old infants were able to rely not only on the kinematic properties of the syllables but also on the newly developed edge-sensitive mechanisms. In fact, it is probably the latter that allowed then to detect the finer detail necessary to discriminate between the two syllables.

The changing nature of object perception mechanisms during infancy may be one factor that accounts for the differential responsiveness to visual as opposed to auditory changes observed in the syllable studies. Another potential factor may be the differential role that visual and auditory information

plays in specifying speech. In studies with adult subjects Massaro and his colleagues [49, 50] have shown that even though visual information contributes in important ways to the perception of speech, it plays a weaker role than does auditory information in the identification of syllables. Specifically, in cases where incongruent audible and visible syllables were presented, the accuracy of visual identification of the syllable was more adversely affected by inconsistent auditory information than the accuracy of auditory identification by inconsistent visual information. Thus, it may be that the failure of the two younger groups of infants to respond to the visible change may have been the result of a combination of their inability to detect the finer detail in the visual modality and the greater power of the auditory information, which in this case indicated that no change occurred.

3.6. The Role of Temporal Synchrony and Featural Information in the Perception of Bimodal Syllables

One of the features of the multimodal component variation method is that the unimodal test trials, by default, produce a temporal desynchronization of the audible and visible components of the face/voice compound stimulus. As a result, there is the possibility that some of the findings described so far might be accounted for by responsiveness to temporal asynchrony *per se* rather than to the change in the features relevant to each study. This possibility, however, can be readily dismissed on logical grounds. If asynchrony were the determining factor then infants would have exhibited significant response recovery in all of the unimodal test trials in all of the studies. Because this was not the case it suggests that other factors played a role either alone or together with the intermodal temporal relations inherent to the face/voice compound stimulus.

Of course, ruling out the possibility that asynchrony alone was not the principal determinant of responsiveness does not mean that temporal relations are not an important factor in intermodal perception. There is certainly substantial evidence indicating that infants are sensitive to intermodal temporal relations. For example, infants have been shown to prefer to look more at bimodal speech whose audible and visible components are in synchrony than when they are out of synchrony [51]. In general, similar to adults [6, 52-54] infants have been found to be sensitive to the temporal relations between audible and visible inputs [13, 14, 55, 56]. Much of the evidence showing that infants are sensitive and responsive to auditory-visual temporal relations, however, comes from studies in which relatively simple non-speech events were

presented (e.g., flashing checkerboards, simple moving objects, and simple percussive sounds). As a result, it is not known currently whether infants can respond to the temporal relations between the visible and audible aspects of bimodal syllables.

To investigate this question, we conducted a study by testing separate groups of 32 infants, aged 4, 6, and 8 months [44]. They were habituated to a talking female face whose vocal and facial features were synchronous while uttering the syllable /ba/. To determine whether the degree of motion might influence responsiveness, one half of the infants in each age group heard and saw the syllable uttered in the AD manner and the other half heard and saw the syllable uttered in the ID manner. Following habituation, infants were given two types of test trials. The first test trial, called the asynchrony familiar-person (AsncFP) test trial, was designed to test responsiveness to asynchrony alone and thus involved the presentation of the familiar talking face whose vocal and facial components were presented 666 ms. out of synchrony. We chose this degree of asynchrony based on the fact that infants need a minimum asynchrony of 350 ms. to detect the asynchrony in a simple A-V compound stimulus [56] and on the assumption that detection of the asynchrony between the audible and visible components of a syllable might be more difficult.

Each infant also was given a second test trial, called the asynchrony novel-person (AsncNP) test trial, during which a novel person whose vocal and facial components also were asynchronous was presented. This test trial was included to determine whether and how the featural characteristics associated with a different person might contribute to responsiveness to syllabic asynchrony. Under normal circumstances 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 of multiple people. Consequently, in addition to having to perceive the temporal synchrony of the A and V components of a speech utterance, they also must process it together with the specific featural characteristics of the person speaking. As a result, the second test trial provided information on the simultaneous processing of these two types of information.

Initial analyses comparing responsiveness as a function of manner of speech indicated that there were no differences. As a result, the data from the two subgroups at each age were combined and these can be seen in Figure 8. As can be seen, the 4-month-old infants did not discriminate the difference between synchrony and asynchrony but did discriminate the change to a novel person. The two older groups exhibited significant discrimination of

the change in the temporal relation as well as in the combined temporal relation/person change.
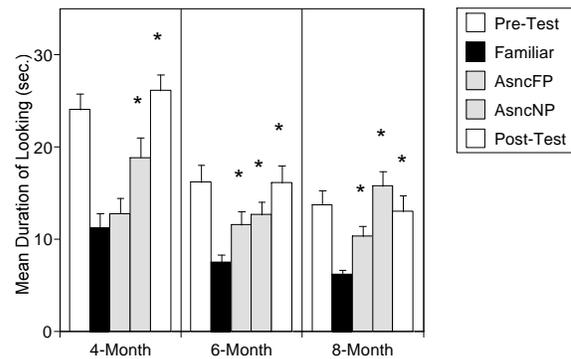

Figure 8. Mean duration of looking at a bimodally asynchronous syllable spoken by a familiar or a novel person following habituation to a synchronous syllable. Asterisks indicate significant response recovery to the type of change indicated by the bar underneath.

To find out how the concurrent availability of the bimodal featural information affected responsiveness, planned comparison tests, contrasting the duration of looking in the AsncFP and AsncNP test trials, were performed. Results indicated that the 4-month-old infants responded significantly more to the AsncNP than to the AsncFP change (p < .001), as did the 8-month-old infants (p < .001). In contrast, the 6-month-old infants did not respond differently to the two types of changes. This suggests an interesting developmental pattern. The absence of a significant response to asynchrony in the AsncFP test trial and a significant response to the asynchrony/feature change in the AsncNP test trial in the 4-month-old infants suggests that the significant response 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. The current study does not make it possible to distinguish between these two alternatives without a test trial in which only a feature change is presented. Prior studies, however, using the multimodal component variation method and presenting talking faces have shown that infants of this age are sensitive and responsive to featural changes alone [31, 40]. As a result, the 4-month-olds most likely responded to the featural change alone in the AsncNP test trial.

The pattern of findings in the 6-month-old infants suggests that they probably based their responsiveness on temporal synchrony relations in both test trials. Prior studies have shown that 6-month-old infants also are sensitive to featural information. If, however, these infants had attended, both, to temporal synchrony relations and featural

information, and if each cue were perceived independently and contributed to responsiveness in some additive fashion then responsiveness in the AsncNP test trial should have been significantly greater than that in the AsncFP test trial. This was not the case. In contrast, the 8-month-old infants did exhibit significantly greater responsiveness in the AsncNP test trial suggesting that they perceived the intermodal temporal relations and the featural information independently. Taken together, the results from this last study show that the capacity to perceive intermodal temporal relations inherent in a bimodally specified syllable emerges sometime between four and six months of age and that the ability to process featural cues independently of temporal relations emerges somewhat later.

4. CONCLUSION

Some might argue that one of the more daunting perceptual tasks that an infant has to perform during the first year of life is the integration of heteromodal sources of information. In order to make sense of the myriad multimodal inputs (both modality-specific and amodal ones) an infant must somehow perceive them as belonging to distinct and unified objects and events. E. Gibson's amodal invariance detection view would suggest, however, that the task is not all that daunting because infants come into the world already prepared to detect a variety of intermodal invariant relations and that they simply get better at doing so as they develop and mature. As a result, according to the amodal invariance detection view, the integration of multimodal inputs is a relatively automatic process. Piaget's theoretical view, on the other hand, would suggest that the task might, in fact, be a daunting one. Indeed, the data reviewed in the current article are consistent with this view.

If amodal invariance detection were the principal mechanism underlying the perception of heteromodal sources of input then infants should have detected and discriminated the unimodal component changes. This is because a unimodal test trial, by definition, involves a disruption of the amodal invariance and should be detected by an organism that is principally tuned to it. The fact that infants in the studies reviewed here did not always do so is inconsistent with the amodal invariance detection view and raises the possibility that other intermodal perceptual processes might be operating at the same time. In other words, the problem of intermodal perception may be one of figure-ground. It may be that in some cases, where the presence of an amodal invariant is highly salient (i.e., where the information given in the stimulus array is relatively simple and non-embedded) the detection of amodal invariance occurs. In other cases, however, where the information in the stimulus array is highly complex because of the concurrent presence of a variety of hierarchically embedded perceptual cues, infants may be unable to detect the invariant relations and thus might need to learn through experience how to integrate the multimodal information.

The latter view is complementary to the amodal invariance detection view but also expands on it by incorporating such developmental principles/facts as the uneven and sequential development of the sensory systems, the critical effects of experience, and the importance of developmental timing effects [5, 57-62]. Needless to say, the studies reviewed here do not specifically provide a test of these developmental principles. Nonetheless, these principles/facts help us to broaden our theoretical lens and thus permit for the possibility that other types of effects besides amodal invariance detection might be expected in early development. Some of the specific ways in which this theoretical broadening might occur is to admit to two possibilities: (1) differential responsiveness in the different sensory modalities, and (2) responsiveness to input in one sensory modality without responsiveness to input in another modality. As a result, it then becomes possible to ask: which sources of information does the infant utilize at a given point in development to perceive a multimodally specified object or event? The latter view also frees investigators from the shackles of amodal invariance detection and permits one to consider other types of intermodal perceptual processes as equally meaningful and informative.

This kind of approach bears some similarities to the approach favored by Massaro in which the contribution of all sources of information is assessed separately based on the assumption that all sources of information may be used by the perceiver to determine the identity of the information. As the data reviewed here make clear, however, in a developing system the evaluation of all available sources of information may not be possible due to developmental limitations (e.g., sensory, perceptual, cognitive). Also, the current data suggest that the different sources of information specifying bimodal speech may not be assessed independently of one another in a developing perceptual system. Finally, because of undeveloped intermodal integration abilities, whatever heteromodal sources are available may not be integrated.

The current approach differs from Massaro's in that it takes development and its unique characteristics into account. It does not as yet have the predictive power that the fuzzy logical model does of adult human performance partly because we are still largely in the "descriptive" phase of the research and partly because it is not possible to use the kinds of experimental methods used with adult subjects with

infants. What the current approach does offer, however, is a way to study the developmental roots of bimodal speech perception and a way to conceptualize it in the context of a developmental system.

5. REFERENCES

1. Burnham, D., "Language specificity in the development of auditory-visual speech perception", in *Hearing by eye II: Advances in the psychology of speechreading and auditory-visual speech*, R. Campbell, B. Dodd, and D. Burnham, Editors, Psychology Press/Erlbaum (Uk) Taylor & Francis: Hove, England UK. p. 27-60, 1998.

2. Aslin, R.N. and L.B. Smith, "Perceptual development", in *Annu Rev Psychol.* p. 435-73, 1988.

3. Banks, M.S. and P. Salapatek, "Infant visual perception", in *Handbook of Child Psychology*, M.M. Haith and J.J. Campos, Editors, John Wiley & Sons: New York. p. 435-571, 1983.

4. Turkewitz, G., J. Gardner, and D.J. Lewkowicz, "Sensory/perceptual functioning during early infancy: The implications of a quantitative basis for responding.", in *Behavioral evolution and integrative levels*, G. Greenberg and E. Tobach, Editors, Lawrence Erlbaum Associates: Hillsdale. p. 167-195, 1984.

5. Gottlieb, G., "Ontogenesis of sensory function in birds and mammals", in *The biopsychology of development*, E. Tobach, L.R. Aronson, and E. Shaw, Editors, Academic Press: New York. p. 67-128, 1971.

6. Massaro, D.W., *Perceiving talking faces: From speech perception to a behavioral principle*. MIT Press/Bradford Books series in cognitive psychology, Cambridge, MA: The MIT Press. xii, 500, 1998.

7. Piaget, J., *The origins of intelligence in children*, New York: International Universities Press., 1952.

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

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

10. Gibson, E.J., "Perceptual development from the ecological approach", in *Advances in Developmental Psychology*, M.E. Lamb, A.L. Brown, and B. Rogoff, Editors, Lawrence Erlbaum Associates: Hillsdale, N.J. p. 243-286, 1984.

11. Lewkowicz, D.J. and G. Turkewitz, "Cross-modal equivalence in early infancy: Auditory-visual intensity matching". *Developmental Psychology*, *16*, 597-607, 1980.

12. Lewkowicz, D.J., "Developmental changes in infants' bisensory response to synchronous durations". *Infant Behavior & Development*, *9*, 335-353, 1986.

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

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

15. Bahrick, L.E. "Increasing specificity in the development of infants' sensitivity to two nested amodal relations". in *International Conference on Infant Studies*. 1996. Providence, RI.

16. Bahrick, L.E., "The development of infants' sensitivity to arbitrary intermodal relations". *Ecological Psychology*, *6*, 111-123, 1994.

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

18. Kuhl, P.K. and A.N. Meltzoff, "The bimodal perception of speech in infancy". *Science*, *218*, 1138-1140, 1982.

19. Lewkowicz, D.J., "Bisensory response to temporal frequency in 4-month-old infants". *Developmental Psychology*, *21*, 306-317, 1985.

20. Lewkowicz, D.J., "Limitations on infants' response to rate-based auditory-visual relations". *Developmental Psychology*, *30*, 880-892, 1994.

21. Lewkowicz, D.J., "Reflections on infants' response to temporally based intersensory equivalence: Response to Spelke (1994)". *Infant Behavior & Development*, *17*, 289-292, 1994.

22. Spelke, E.S., "Preferential looking and intermodal perception in infancy: Comment on Lewkowicz (1992b)". *Infant Behavior & Development*, *17*, 284-286, 1994.

23. Welch, R.B. and D.H. Warren, "Intersensory interactions", in *Handbook of perception and human performance: Sensory processes and perception*, K.R. Boff, L. Kaufman, and J.P. Thomas, Editors, J. Wiley & Sons: New York. p. 1-36, 1986.

24. McGurk, H. and J. MacDonald, "Hearing lips and seeing voices". *Nature*, *264*, 229-239, 1976.

25. Radeau, M. and P. Bertelson, "Adaptation to auditory-visual discordance and ventriloquism in semirealistic situations". *Perception & Psychophysics*, *22*, 137-146, 1977.

26. Radeau, M. and P. Bertelson, "Cognitive factors and adaptation to auditory-visual discordance". *Perception & Psychophysics*, *23*, 341-343, 1978.

27. Radeau, M., "The locus of adaptation to auditory-visual conflict". *Perception*, *2*, 327-332, 1973.

28. Dodd, B. and D. Burnham, "Processing speechread information". *Volta Review*, *90*, 45-60, 1988.

29. Walker-Andrews, A.S. and E. Lennon, "Infants' discrimination of vocal expressions: Contributions of auditory and visual information". *Infant Behavior & Development*, *14*, 131-142, 1991.

30. Rosenblum, L.D., M.A. Schmuckler, and J.A. Johnson, "The McGurk effect in infants". *Perception & Psychophysics*, *59*, 347-357, 1997.

31. Lewkowicz, D.J., "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, 1996.

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

33. Cooper, R.P. and R.N. Aslin, "Preference for infant-directed speech in the first month after birth". *Child Development*, *61*, 1584-1595, 1990.

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

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

36. Werker, J.F., J.E. Pegg, and P.J. McLeod, "A cross-language investigation of infant preference for infant-directed communication". *Infant Behavior & Development*, *17*, 323-333, 1994.

37. Pegg, J., J.F. Werker, and P.J. McLeod, "Preference for infant-directed over adult-directed speech: Evidence from 7-week-old infants". *Infant Behavior and Development*, *15*, 325-345, 1992.

38. DeCasper, A. and M.J. Spence, "Auditorily mediated behavior during the perinatal period: A cognitive view", in *Newborn attention: Biological constraints and the influence of experience*, M.J.S. Weiss and P.R. Zelazo, Editors, Ablex Publishing Corp.: Norwood, N.J., 1991.

39. Trehub, S.E., A.M. Unyk, and L.J. Trainor, "Maternal singing in cross-cultural perspective". *Infant Behavior & Development*, *16*, 285-295, 1993.

40. Lewkowicz, D.J., "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, 1998.

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

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

43. Jusczyk, P.W., D. Houston, and M. Goodman, "Speech perception during the first year", in *Perceptual Development: Visual, Auditory, and Speech Perception in Infancy*, A. Slater, Editor, Psychology Press: East Sussex, UK. p. 357-388, 1998.

44. Lewkowicz, D.J., "Infants' response to the audible and visible properties of the human face: Perception of syllable differences". submitted.

45. Dannemiller, J.L. and R.L. Freedland, "Detection of relative motion by human infants". *Developmental Psychology*, *27*, 67-78, 1991.

46. Roessler, J.S. and J.L. Dannemiller, "Changes in human infants' sensitivity to slow displacements over the first 6 months". *Vision Research*, *37*, 417-423, 1997.

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

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

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

50. Massaro, D.W. and M.M. Cohen, "Perceiving speech from inverted faces.". *Perception and Psychophysics*, *58*, 1047-1065, 1996.

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

52. Dixon, N.F. and L.T. Spitz, "The detection of auditory visual desynchrony". *Perception*, *9*, 719-721, 1980.

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

54. Radeau, M., "Auditory-visual spatial interaction and modularity". *Cahiers de Psychologie*, *13*, 3-51, 1994.

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

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

57. Gottlieb, G., "Experiential canalization of behavioral development: Theory". *Developmental Psychology*, *27*, 35-39, 1991.

58. Gottlieb, G., *Synthesizing Nature-Nurture: Prenatal Roots of Instinctive Behavior*, Mahwah, NJ: Erlbaum, 1997.

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

60. Turkewitz, G. and P.A. Kenny, "Limitations on input as a basis for neural organization and perceptual development: A preliminary theoretical statement". *Developmental Psychobiology*, *15*, 357-368, 1982.

61. Turkewitz, G. and P.A. Kenny, "The role of developmental limitations of sensory input on sensory/perceptual organization". *Journal of Developmental and Behavioral Pediatrics*, *6*, 302-306, 1985.

62. Turkewitz, G., "Sources of order for intersensory functioning", in *The development of intersensory perception: Comparative perspectives*, D.J. Lewkowicz and R. Lickliter, Editors, L. Erlbaum: Hillsdale. p. 3-17, 1994.