AN INVESTIGATION OF YOUNG INFANTS' ABILITY TO MATCH PHONETIC AND GENDER INFORMATION IN DYNAMIC FACES AND VOICES

by

MICHELLE LOUISE PATTERSON

B.A. (Honours), Queen's University, 1996
M.A. The University of British Columbia, 1998

A THESIS SUBMITTED IN PARTIAL FULFILLMENT OF
THE REQUIREMENTS FOR THE DEGREE OF

DOCTOR OF PHILOSOPHY

in

THE FACULTY OF GRADUATE STUDIES

Department of Psychology

We accept this thesis as conforming
To the required standard

THE UNIVERSITY OF BRITISH COLUMBIA

May 2001

In presenting this thesis in partial fulfilment of the requirements for an advanced degree at the University of British Columbia, I agree that the Library shall make it freely available for reference and study. I further agree that permission for extensive copying of this thesis for scholarly purposes may be granted by the head of my department or by his or her representatives. It is understood that copying or publication of this thesis for financial gain shall not be allowed without my written permission.

Department of PSYCHOLOGY

The University of British Columbia
Vancouver, Canada

Date June 24, 2002

DE-6 (2/88)

Abstract

This dissertation explores the nature and ontogeny of infants' ability to match phonetic information in comparison to non-speech information in the face and voice. Previous research shows that infants' ability to match phonetic information in face and voice is robust at 4.5 months of age (e.g., Kuhl & Meltzoff, 1982; 1984; 1988; Patterson & Werker, 1999). These findings support claims that young infants can perceive structural correspondences between audio and visual aspects of phonetic input and that speech is represented amodally. It remains unclear, however, specifically what factors allow speech to be perceived amodally and whether the intermodal perception of other aspects of face and voice is like that of speech. Gender is another biologically significant cue that is available in both the face and voice. In this dissertation, nine experiments examine infants' ability to match phonetic and gender information with dynamic faces and voices.

Infants were seated in front of two side-by-side video monitors which displayed filmed images of a female or male face, each articulating a vowel sound (/a/ or /i/) in synchrony. The sound was played through a central speaker and corresponded with one of the displays but was synchronous with both. In Experiment 1, 4.5-month-old infants did not look preferentially at the face that matched the gender of the heard voice when presented with the same stimuli that produced a robust phonetic matching effect. In Experiments 2 through 4, vowel and gender information were placed in conflict to determine the relative contribution of each in infants' ability to match bimodal

information in the face and voice. The age at which infants do match gender information with my stimuli was determined in Experiments 5 and 6. In order to explore whether matching phonetic information in face and voice is based on featural or configural information, two experiments examined infants' ability to match phonetic information using inverted faces (Experiment 7) and upright faces with inverted mouths (Experiment 8). Finally, Experiment 9 extended the phonetic matching effect to 2-month-old infants. The experiments in this dissertation provide evidence that, at 4.5 months of age, infants are more likely to attend to phonetic information in the face and voice than to gender information. Phonetic information may have a special salience and/or unity that is not apparent in similar but non-phonetic events. The findings are discussed in relation to key theories of perceptual development.

TABLE OF CONTENTS

An Investigation of Young Infants' Ability to Match Phonetic and Gender Information in Dynamic Faces and Voices

LITERATURE REVIEW

In general, our ability to detect intersensory relations reflects the integrative activity of the central nervous system and forms the foundation upon which many higher-level perceptual, cognitive, and linguistic functions are based (Lewkowicz, 1999). Although questions about the development of intersensory abilities have made a significant contribution to discussions of perceptual development, it is only over the past two decades that researchers have begun to probe infants' intersensory abilities. Infancy provides the opportunity to study the ontogenetic development of intersensory functioning and reveals such abilities in rudimentary forms.

Although infants have to learn most intermodal relations, young infants have been shown to be skilled perceivers of *amodal invariants* (Bahrick, 1988, 1992; Spelke, 1976; Walker-Andrews, 1994). Amodal invariants are perceptual cues that are tied to the structural properties of an action or event and are not specific to a particular sensory modality but are redundant across two or more modalities. For example, the sight and sound of hands clapping share temporal synchrony, a common tempo of action, and a common rhythm. The amodal properties of synchrony, tempo, and rhythm are concurrently available both visually and acoustically. Young infants are remarkably adept at matching audio-visual events based on rhythm, tempo, and even affective and

phonetic information in the face and voice (see Lewkowicz, 2000; Bahrick, 2000; Bahrick & Pickens, 1994; Kellman & Arterberry, 1998 for reviews).

Audio-visual speech perception may be a special case of intermodal perception. It has been suggested that speech may engage distinct sensorimotor perceptual mechanisms that render articulation and heard speech equivalent (Liberman & Mattingly, 1985)[1]. A specialized speech processing mechanism would allow infants to perceive speech as an amodal event, perhaps enabling intermodal matching at an early age (Kuhl & Meltzoff, 1982; 1984; MacKain, Studdert-Kennedy, Spieker, & Stern, 1983). Recent research (Patterson & Werker, 1999) with both female and male faces replicates and extends past findings that infants at 4.5 months of age can detect matching phonetic information in the face and voice. These findings are consistent with the hypothesis that young infants' ability to match phonetic information in the lips and voice may be based on the detection of amodal invariants.

The age at which infants are first able to detect amodal invariants for speech is still an open question as is the relationship between detecting speech versus other nonspeech events, such as gender information, in faces and voices. Faces and voices have been described as thoroughly intermodal "objects of perception" (Kuhl & Meltzoff, 1988) because the same information (e.g., speech, affect, gender) can be

[1] Proponents of the motor theory of speech perception claim that phonetic intentions of the speaker are represented in a specific form in the speaker's brain and that there is a perceiving module specialized to lead the listener effortlessly to that representation. The fact that young infants seem to detect audio and visual aspects of vowel articulations relatively easily and also engage in some articulatory imitation provides some support for motor theory's assumption that the phonetic mode, and the perception-production link it incorporates, is innately specified. Conclusive evidence to test this hypothesis would require conducting a similar study with newborns; however, our procedure is not suitable for newborns thus support for this explanation awaits further methodological advancements.

detected through sight, sound, and often touch. Thus, faces and voices can be used to examine central issues in theories of intermodal perception such as how audio and visual information are organized to form a unified percept and whether there is a common metric that recognizes the equivalence between information entering different sensory channels. A central question this dissertation addresses is whether all audio-visual information in the face and voice is equivalent. Specifically, a series of nine studies examines when and how infants match phonetic and gender information in faces and voices as well as how these two sources of information interact in audio-visual perception.

Traditional Theories of Intersensory Development

For many years theorists have been interested in how information that is processed by separate modalities is equated in perception. Two prevailing theoretical views, known respectively as the integration view and the differentiation view, have dominated the study of intersensory development until recent years (see Lickliter & Bahrick, 2000; Lewkowicz, 2000 for reviews). The *integration view* holds that the different sensory modalities function as separate sensory systems during the initial stages of postnatal development and gradually become integrated during development through the infant's activity and repeated experience with concurrent information provided by the different sensory modalities (Piaget, 1952; Birch & Lefford, 1963). Until integration of the senses is achieved, infants presumably perceive unrelated patterns of light and sound from a single multimodal event.

In contrast, the *differentiation view* holds that the different senses form a primitive unity early in development and that young infants are able to relate concomitant input to different modalities by detecting "amodal invariants" (Gibson, 1969). As the infant develops, the modalities differentiate from one another and the information arising from the different sensory modalities becomes further differentiated as well. Infants are thought to differentiate progressively finer and more complex multimodal relations through their experience over the course of development (Gibson, 1969; Bower, 1974)[2]. If one takes this view, it is essential to determine what information infants are sensitive to as well as the limits and mechanisms underlying such sensitivity (Walker-Andrews, 1994).

As a result of these opposing views, the most prominent questions guiding research on early intersensory development over the past several decades have tended to focus on (a) whether intersensory development proceeds from initially separate senses that become increasingly integrated through experience, eventually resulting in coordinated multimodal perception, or (b) whether the development of intersensory perception is a process of differentiation and increasing specificity.

Although some controversy remains as to whether perceptual development proceeds according to the integration or the differentiation view, the current consensus argues against what now appears to be an artificial dichotomy between integration and

---

[2] Audiovisual speech perception has also been cited as support for the Gibsonian/differentiation view of perceptual development. The Gibsonian view is similar to motor theory in that both propose that speech perception involves recovering the articulatory gesture, however, the Gibsonians do not think that speech perception requires a special processing module (e.g., Fowler & Rosenblum, 1991). Articulatory gestures are distal events which, through lawful relationships, give rise to a proximal stimulus—the acoustic signal and the visual signal. Thus, evidence that audiovisual integration is a mandatory property of perception can be appreciated without reference to modularity (Fowler & Rosenblum, 1991). Speech perception is like the perception of any other distal event.

differentiation processes (Kellman & Arterberry, 1998). Increasing focus on infant intersensory perception over the past few decades has provided mounting evidence that both processes are involved in perceptual development. On the one hand, there is now compelling neuroanatomical, neurophysiological, and behavioral evidence of strong intermodal linkages in newborns and young infants from a variety of species, including humans (Lewkowicz & Turkewitz, 1980). For example, infant animals are more likely to show sensitivity to intersensory correspondences than are older animals in a classical-conditioning learning paradigm (Spear & McKinzie, 1994), and human infants demonstrate an array of intermodal perceptual skills in the weeks and months following birth, including intersensory facilitation, in which stimulation in one modality enhances responsiveness to stimuli in other modalities (Lewkowicz & Lickliter, 1994). It has been shown that by 4 months of age infants can detect many *amodal invariants* even if they have had little or no experience with the relations in question. Given that the role of integration is presumably to link together separate sensations to form a unified percept, in the common cases where redundant information is obtained by two or more perceptual systems, infants' detection and abstraction of amodal relations makes the need for intersensory integration unlikely (Lickliter & Bahrick, 2000).

On the other hand, infants can also detect perceptual information that is not amodal in the months following birth and are not able to detect some amodal relations (e.g., rate-based intersensory equivalence without synchrony) until the end of the first year of life (e.g., Humphrey & Tees, 1980; Lewkowicz, 1985; 2000). Many relations

between stimulus properties conveyed concurrently to different modalities are arbitrary in the sense that they are not united by information common across the different sensory modalities and can vary as a function of context or stimulus domain. For example, the relation between a person's appearance and the specific sound of his or her voice is an arbitrary pairing of stimulation across two or more sensory modalities. Given that there is no common information that links the stimulation presented to the two or more modalities, such arbitrary relations must be learned by experiencing the information in the modalities together and may thus be characterized as depending on some process of integration.

Although the interaction between differentiation and integration and their complementary roles in the emergence and maintenance of intersensory functioning are worthy of more systematic investigation, it has been argued that this framework alone will prove inadequate in forging a more complete and coherent view of early intersensory perception (Lickliter & Bahrick, 2000). For example, both views tend to under-characterize the complex and dynamic processes of organization and reorganization within and between sensory systems documented in the study of animal infants' prenatal and postnatal development (Banker & Lickliter, 1993; Symons & Tees, 1990). Although the primary difference between the two perspectives concerns the developmental process, few studies of intermodal perception are developmental in nature; most developmental trends are drawn from several single age group studies. Furthermore, the most robust findings are observed with infants aged 4-months and older and by this age the effects of nature and nurture are already tightly entwined.

This is particularly pertinent to speech perception since the infant has been exposed to maternal vocalizations prenatally both by air and by bone conduction (Cooper & Aslin, 1989) and will have had at least some experience producing approximations to speech sounds in pre-babbling vocalizations (Oller, 1986). Indeed, several common assumptions underlying the integration and differentiation perspectives of early human development have been called into question (see Lickliter & Bahrick, 2000).

More recent theories of perceptual development reflect the blurred boundary between nature and nurture and emphasize epigenetic (Lewkowicz, 2000; Werker & Tees, 1992) or innately guided learning (Jusczyk & Bertoncini, 1988) processes as important for intermodal perception. In this light, recent research is beginning to clarify how the processes of differentiating amodal relations and integrating arbitrary or modality-specific information across the senses can interact with one another in certain domains or in the development of certain skills. For example, Lewkowicz (2000) argues that his research on infants' ability to discriminate auditory, visual, and bimodal changes in talking faces supports an epigenetic-systems view of intermodal development. Lewkowicz argues that perception need not be amodal from birth; development is a system of epigenetic interactions where structural and functional limitations determine a sequential hierarchal emergence of responsiveness. The co-action of differentiation and integration in addition to interactions of factors both intrinsic and extrinsic to the organism propels development.

In this dissertation, the literature will be reviewed in three sections; however, before turning to this review, I will briefly outline what is known about infants' ability

to perceive audio-visual equivalents in objects and events as well as their ability to link more general attributes in faces and voices.

Audio-Visual Perception of Objects and Events

Research on infant audio-visual perception has largely been conducted from a differentiation (i.e., Gibsonian) framework. A large body of research has consistently demonstrated a number of principles that are now important cornerstones of Gibsonian intermodal theory. Since, by 4 months of age, infants can relate a variety of audio-visual concomitants, Gibsonians argue that intermodal development is set in motion and guided by the detection of amodal invariant relations and occurs in order of increasing specificity. This claim is supported by findings that global amodal relations are detected developmentally prior to nested amodal relations. Secondly, research from several domains demonstrates that many amodal relations are generally detected developmentally prior to arbitrary relations. Third, evidence suggests that detection of amodal relations guides and constrains perceptual learning about arbitrary relations. This section will briefly review evidence of these findings (see Bahrick, 2000; Kellman & Arterberry, 1998; and Lickliter & Bahrick, 2000 for reviews). The primary experimental methods that have been used to study infant intersensory development are described in Appendix A.

As mentioned, research has shown that intermodal matching improves with age. Infants detect global amodal relations (e.g., temporal synchrony, rhythm) prior to nested amodal relations (e.g., object composition; Bahrick, 1983; 1987; Lewkowicz, 1985, 1992). This pattern appears to be adaptive. By first detecting temporal synchrony, for

example, infants can focus on unitary events and further differentiation will be appropriately constrained. The initial focus on global, synchrony relations creates a natural buffer against processing unrelated streams of visual and acoustic stimulation. By ensuring that attention is first focused on audible and visible stimulation that belong together, further processing of multimodal events can proceed in an economical and veridical way.

Prior to 3 months of age, infants are already sensitive to many amodal relations (e.g., temporal synchrony, tempo, rhythm), but they are not able to detect arbitrary relations (e.g., pitch with colour or shape) until 7 to 9 months of age (Bahrick, 1994). This suggests that there may be a developmental lag between the detection of amodal and arbitrary relations from a given set of events. Although amodal relations can be directly perceived and are context-free, arbitrary relations must be learned and may vary from one context or event to another. Young infants are able to learn arbitrary relationships between auditory and visual information, particularly if auditory and visual information are spatially coincident and temporally synchronous (e.g., Lyons-Ruth, 1977; Morrongiello et al., 1998). It appears that the detection of amodal relations developmentally precedes and constrains the detection of arbitrary relations in a given domain.

Much of the work influenced by the Gibsonian view has used natural, ecologically-valid stimuli. In an attempt to determine exactly what properties are guiding intermodal perception, other researchers have used simple, artificial stimuli such as lights and tones that lack ecological validity but control for many confounding

variables (e.g., Lewkowicz, 1985; Humphrey, Tees, & Werker, 1979). Such stimuli may not be as revealing as more natural stimuli for understanding intermodal development in infants. Conflicting findings in the literature are likely due to different experimental paradigms, procedures, and the nature of the stimuli. In general, more robust evidence of intermodal matching has been found when stimuli are ecologically valid. It is possible that in a more complete informational context, the presence of multiple modality-specific cues might help infants differentiate two visual stimuli intramodally and then associate them intermodally (Lewkowicz, 1992).

In summary, infants appear to detect intersensory relations in a particular order developmentally. When multimodal events make both amodal and arbitrary relations available, as is typical in the natural environment, infants generally first differentiate global amodal synchrony relations. Later, they differentiate nested amodal relations such as information specifying aspects of the object's visual appearance and its sound or touch. There appears to be a developmental lag between the detection of global amodal, nested amodal, and arbitrary relations within a given domain or set of events.

Audio-Visual Perception of Faces and Voices

Faces and voices are arguably the most important signals in the human visual and auditory domains. The mechanisms involved in relating audio-visual information about the face and voice may well differ from those involved in relating audio-visual aspects of other objects and events. It may be that there are even different kinds of relationships between faces and voices. Some face-voice relations may be *arbitrary but natural* (e.g., particular face and voice), some may be either arbitrary but natural or

amodal (e.g., gender of face and voice), and others may be more difficult to define (Walker-Andrews, 1994). For example, some controversy exists over whether the relation between phonemes and articulatory movements is learned, amodal, or something more like innately-guided learning. Whatever the relationship(s) between various aspects of face and voice, there is something qualitatively different in learning about these stimuli in comparison to learning about objects and sounds. What is not as clear is whether learning about face-voice relations is different for speech tasks versus other nonspeech face-voice relations. Few researchers have examined what aspects of human speech affect the perception of equivalent audio and visual information. The studies reviewed in this section used natural faces and voices to study infants' general ability to connect these biologically relevant stimuli as well as the basis and the ecological implications for such a connection.

General Characteristics of Face and Voice

During the first year of life, infants become increasingly sensitive to their parents' faces and voices and the relation between them. It is well known that newborns prefer their mother's voice over a female stranger's voice (DeCasper & Fifer, 1980) and prefer faces over non-face stimuli (see Nelson & Ludemann, 1989 for a review). What is not as well understood is how infants *relate* faces and voices. Cohen (1974) found that 8-month-olds, but not 5-month-olds, looked longer when the face and voice of their mother and a female stranger belonged together than when the face and voice did not match. Spelke and Owsley (1979) reported that infants between 3.5- and 7.5-months preferred to look at a motionless parent whose voice was played over a speaker rather

than at a moving parent whose voice was not played, despite the absence of any face-voice synchrony. Results were more pronounced as infants got older. When the father was replaced by a female stranger, infants preferred to look at the mother when the stranger's voice was played and vice versa. These studies suggest that infants are able to process audio and visual attributes somewhat independently and associate them when the attributes are ecologically related.

In this dissertation, Section I examines whether gender matching emerges at the same time as phonetic matching. Section II explores some possible bases of phonetic matching in infants and, finally, Section III examines the phonetic matching effect in 2-month-old infants.

SECTION I

MATCHING GENDER INFORMATION IN DYNAMIC FACES AND VOICES

Section I explores whether young infants show the same ability to match information about gender in the face and voice as they do bimodal phonetic information. Gender is a salient and naturally occurring class to which infants typically have been exposed. Invariant gender information in the face and voice has been described as a *natural and typical* multimodal (rather than amodal) relation (Walker-Andrews, 1994). Natural and typical relations are described as frequently occurring but are neither exclusively amodal nor exclusively arbitrary. They may be comprised of sets of overlapping relations, some inherent in audio-visual structure (i.e. amodal) and some that require a period of learning (i.e. arbitrary). As such, infants may not be able

to match on the basis of this type of relation until an older age than shown for matching of amodal relations.

The classification of gender information as *natural and typical* may depend upon the kind of gender information to which the infant is exposed. Visually, gender can be differentiated on the basis of physiognomic properties such as size of features, especially the nose (Chronicle et al., 1995), and skin texture (Brown & Perret, 1993) as well as culture-specific variations such as facial hair, hair length and make-up. In terms of voices, men's tend to be lower pitched, although there is overlap. In addition, vowels produced by a man, woman, or child are different in terms of formant frequency (Peterson & Barney, 1952). Such differences result primarily from the length of the vocal tract and the size of the vocal folds (Ladefoged, 1993) and thus may have visible correlates. For example, the human male adult tends to be larger than the female; he has a longer and bigger throat, a more protuberant Adam's apple, a larger face, and broader shoulders. Culture-specific cues to gender tend to vary more than physiognomic cues and likely require a longer period of learning (Leinbach & Fagot, 1993). Gender-specific cultural conventions for hair length, hairstyle, clothing, and jewelry (e.g., ear and nose rings) vary tremendously across cultures and even change across time within a culture. Those gender attributes in the face and voice that are primarily physiognomic could be seen as another set of *amodal* relations (like phonetic relations) and thus may be detected by infants younger than 6 months of age. However, to the extent that gender relations are based on culture-specific cues, the

natural and typical classification would be more apt and would predict later detection than most amodal invariants.

Gender categorization is of central importance to human social functioning. Thus, it is possible that evolution may have predisposed infants to have an initial perceptual bias to attend to the physiognomic features of gender. Moreover, the physiognomic characteristics of gender may uniquely specify the sound properties of the heard voice. On the other hand, the enormous degree of cultural variability in the way gender is expressed argues for the advantage of a more open developmental program, allowing infants to learn gradually over the first year of life to associate modality-specific gender information and then generalize to gender categories. A brief review of what is known about the perception and categorization of gender information might help clarify just what kind of information young infants can detect.

Detection and Discrimination of Male/Female Faces and Voices

Habituation and visual preference studies have shown that infants well under one year of age can discriminate male and female faces (Cornell, 1974; Fagan, 1976, 1979). Fagan (1979) used a visual recognition task to demonstrate that 5- and 6-month-olds are more likely to recognize whether or not a face is familiar on the basis of gender than on difference or similarity of facial structure. More recently, in a series of studies with 4-, 6-, and 8-month-olds, Lewkowicz (1996) found evidence that infants are sensitive to the visual cues in the face that distinguish gender. However, compared to the older infants, 4-month-olds were not as able to discriminate stimuli on the basis of only an auditory change in gender. When Lewkowicz (1998) habituated 3-, 4-, 6-, and 8-

month-olds to a male face reciting a script in adult-directed speech and then tested infants' ability to discriminate a change to a woman singing the same script, all age groups (even the 3-month-olds) were able to discriminate visual, auditory, and bimodal changes.

Lewkowicz (1999) argues that, as more features are added, discrimination may occur earlier and to more component changes. However, it is not clear from the discrimination studies just what cues to gender infants are using. In a discrimination study we do not know whether it is the gender of the face and voice or simply a change to a new face/voice that infants detect. Nevertheless, these studies reveal that under at least some circumstances, infants 4 months of age and younger are able to discriminate male from female faces and male from female voices. The ability to discriminate different-gender faces and voices is required for infants to match faces and voices based on gender.

Categorization of Faces and Voices based on Gender

One could argue that, in order to match gender in the face and voice, infants need to not only discriminate these cues but also to categorize on their basis. It may only be with this level of knowledge that gender matching is possible. However, it is only toward the end of the first year of life that infants reliably categorize on the basis of gender. Leinbach and Fagot (1993) assessed categorical responding to pictures of male and female faces (including neck and shoulders) using an infant-controlled habituation procedure. The pictures differed in terms of facial orientation, coloring, hair length, expression, clothing, and presence or absence of facial hair. Infants aged 9- and 12-

months consistently showed evidence of categorical responding but the 5- and 7-month-olds did not. In a follow-up study, different groups of 12-month-olds were habituated to one of the original male and female displays (the male had short hair and wore a shirt and tie while the female had shoulder-length hair and wore a blouse). After habituation, infants were presented with one of four test combinations: (1) the original male and female but both wearing unisex clothing, (2) the female with short hair but original clothing and the original male, (3) the female with short hair and both models wearing unisex clothing, or (4) a control condition (i.e., the habituation stimuli). Infants recovered attention to the first two test conditions, however, they did not recover attention to the control trial or when culture-specific cues to both hair and clothing were removed. This suggests that, by one year of age, infants can form categories for men and women but that they are using cultural-specific cues about sex-typical hair length and clothing style.

Categorization of gender-typical versus gender-ambiguous faces has also been investigated in young infants. Younger and Fearing (1999) habituated 7- and 10-month-old infants to randomized sequences of male and female faces. The 7-month-olds were not able to categorize on the basis of gender. The 10-month-olds did show evidence of categorical responding, but only for faces that fit clearly within the gender-typical category. These results suggest that by 10 months of age infants are sensitive to prototypical information specifying gender categories.

Categorical responding to male and female voices seems to occur earlier than to different-gender faces. Miller (1983) reported that 6-month-olds who heard and

habituated to male or female voices saying "hi" then detected a shift to voices of the opposite sex uttering the same word. Thus, young infants appear to demonstrate gender-based categorical responding to voices by 6 months of age.

Taken together, studies of discrimination and categorization suggest that infants are able to notice gender cues by 3 to 5 months of age, but may not be able to use both heard and seen cues to categorize based on gender until the second half of the first year of life.

Matching Gender Information in Face and Voice

To date, only a handful of studies have explored bimodal gender matching and the results are somewhat contradictory. A few studies have assessed infants' ability to relate static faces and taped voices for male and female stimuli. Spelke and Owsley (1979) tested infants in an intermodal preference procedure using the faces and voices of infants' mothers and fathers. Infants aged 3.5 months did look preferentially at the parent whose voice was heard; however, the effect was more pronounced at 7.5 months of age. It is not clear whether infants were using information related to gender or whether intermodal matching was based on idiosyncratic information learned through experience with particular caregivers. To address this issue, Miller and Horowitz (1980) presented 8-month-olds with paired slides of unfamiliar male and female faces along with male or female taped voices. Infants tended to look longer at the face accompanied by the gender-appropriate voice, but only for the male faces and voices. Lasky, Klein, and Martinez (1974) tested 5-and 6-month-olds with paired photographs of man - woman, man - boy, or woman - boy, accompanied by each voice of the pair in

sequence. Infants looked longer at the male photograph overall and showed a visual preference for the voice-appropriate photographs but only when the woman's face was paired with the boy's face. Therefore, infants may have been responding to the dimension of age (adult versus child) rather than to gender category. In sum, these findings provide little conclusive evidence for intermodal knowledge of faces and voices based on gender cues.

Poulin-Dubois, Serbin, Kenyon, and Derbyshire (1994) investigated infants' intermodal knowledge of gender at 9 and 12 months of age. Infants were presented with side-by-side still photographs of a male and a female face. The male or female voice, reciting a short greeting in infant-directed speech, was played from a centrally-located speaker. Infants at both ages spent more time looking at the picture that matched the heard voice, but only for the female stimuli. Even when highly stereotypical faces and voices were used, preference emerged only for the female stimuli. No preference for either the female or male face independent of sound was found. This work suggests that intermodal knowledge about gender does not emerge until at least the end of the first year. However, all studies described above used static faces accompanied by a voice, which precluded intermodal matching on the basis of dynamic relations and thus may have forced infants to rely on well-learned cultural information.

Little is known about when and under what conditions infants are able to relate dynamic faces and voices based on gender cues. Francis and McCroy (1983, poster cited in Walker-Andrews, Bahrick, Raglioni, & Diaz, 1991) investigated infants' sensitivity to

dynamic, bimodal presentations of male and female adults. Infants at 3-, 6-, and 9-months of age were shown a male and a female face accompanied by a female voice, a male voice, or music. Results revealed that 6-month-olds who heard the male voices looked longer at the male faces compared to infants who heard music or female voices. Infants who heard the female voices looked longer at the female faces compared to infants who heard male voices but not significantly more than infants who heard music. In contrast, the 3-month-olds looked longer to the female face regardless of the auditory condition, whereas the 9-month-olds exhibited no significant looking preferences at all. These mixed results allowed no firm conclusions. Moreover, the fact that two separate video monitors were used for the visual presentations makes it unlikely that the two faces and accompanying voice could have been precisely synchronized throughout. Thus, the role played by face-voice synchrony is not known.

Only one set of studies has examined infants' ability to match gender using dynamic faces and voices. Walker-Andrews et al. (1991) showed infants videos of a male and a female face speaking side-by-side with a single soundtrack (a nursery rhyme) that corresponded to the gender of one visual display but was synchronized with the motions of both displays. In the first of two studies, 4- and 6-month-olds were presented with two 2-min trials where the male and female voices were played for two minutes in a randomized order. The 6-month-olds looked longer at the matching gender display across both trials and the 4-month-olds showed evidence of matching only on the second trial. In the second study, 3.5- and 6.5-month-olds were presented with two blocks of eight 20-sec trials. The 6.5-month-olds showed evidence of matching

but the 3.5-month-olds did not. This study, in contrast to the work with static faces, would suggest that the ability to match dynamic faces and voices based on gender cues may be evident as early as 6 months of age.

In summary, research using various paradigms has found that infants can discriminate male and female faces and voices long before their first birthday. Moreover, young infants have formed gender-based categories. Infants can categorize voices based on gender by 6 months of age and faces by 6- to 10-months of age. Matching faces and voices based on gender seems to develop later and depends on whether faces are static or dynamic. Infants do not look preferentially to gender-appropriate pictures of static faces until 9- to 12-months of age, and even at those ages seem better able to match the face and voice of females, with whom they have perhaps had more exposure. Only one study has examined infants' ability to match dynamic faces and voices based on gender. That study (Walker-Andrews et al., 1991) suggests that infants may be able to match gender in the face and voice by 6 months of age. It is on the basis of these results that the argument is made that gender matching requires *natural and typical* information.

Section I of this dissertation explores whether the age at which infants match dynamic gender information in face and voice will change if the facial and vocal cues to gender are more like those that have been used in studies of phonetic matching. In this dissertation, infants were presented with the same stimuli used in Patterson and Werker's (1999) vowel matching study. Perhaps, by showing infants dynamic displays of faces and voices repeating only isolated vowels, the culture-specific cues to gender in

the voice will be minimized and the attributes of the voice that are more reliant on physical differences between males and females will be more evident. Similarly, by using dynamic displays of male and female speakers, in which some culture-specific visual information (i.e., clothing, make-up, jewelry) is not as readily available, infants may pay more attention to physiognomic features of face and voice. In this way, the current studies may allow direct comparison of infants' ability to match gender information to their ability to match phonetic information in face and voice. If infants can match gender using amodal information in the face and voice then gender matching, like phonetic matching, may be evident at 4.5 months of age. If gender matching is not evident at 4.5 months, such a finding would support the argument that gender information is not entirely amodal and requires a period of learning. In addition, the age at which robust gender matching does appear can be determined.

Speech Information as Intermodal Cues

Though normally regarded as a purely auditory phenomenon, speech perception appears to be profoundly influenced by visual, and even tactual[3], information. Despite a strong visual influence, all consonants and vowels cannot be identified by vision alone. The component features of consonants can be divided into three types: (1) place of articulation describes the location in the vocal tract where the primary constriction of airflow occurs; (2) manner of articulation refers to basic ways articulation can be accomplished and distinguishes phonetic segments produced at the same place of articulation; and (3) voicing refers to the state of the glottis during articulation. Place of

articulation for frontal consonants can be identified visually, but manner and voicing cannot; vowels, on the other hand, can be identified by consistent visual, as well as acoustic, information. English vowels are fairly discriminable based on two properties: (1) the extent of vertical lip separation and tongue height, and (2) the degree of lip spreading or rounding. Although it has been stated that no direct manifestation of intonation or linguistic stress can be gleaned through the visual channel (Ladefoged, 1993), recent research suggests that even intonation may be recovered from global head movements (Vatikiotis-Bateson, 2001).

During typical conversations we see the talker's face and watch the movements of lips, tongue, and jaw that are concomitant by-products of speech. For adults, speech perception is an intermodal phenomenon (for the approximately 33% of phonemes that have visual concomitants). The role of vision in speech perception was fully recognized after the demonstration of the McGurk effect (McGurk & MacDonald, 1976) which places auditory and visual information in conflict. When the audio signal /ba/ is presented with the visual lip movements for /ga/, the illusory percept /da/ is perceived. Recent research has shown that 4-month-old infants appear to integrate audiovisual speech to produce an emergent perception that is not provided either acoustically or visually (Rosenblum et al, 1997; Burnham & Dodd, 1997); however, research by Desjardins (1997) suggests that this ability in 4-month-olds may still be fragile.

---

[3] Some speech information can be delivered to the skin and integrated with speech information perceived by eye or by ear; however, the effect of tactual speech information is not as dramatic as that obtained through vision and requires training (Grant et al., 1986).

A new and complex issue in speech perception research is how knowledge of the intermodal nature of speech is acquired by infants. For example, how is information from different modalities organized to form a unified speech percept? Is there an amodal metric that recognizes the equivalence between information entering the different channels? If so, what is the nature of this common metric; is it in the form of phonetic information or some other amodal specification? Alternatively, perhaps intermodal speech perception relies on an "arbitrary but natural" pairing that must be learned. If so, when and under what conditions do infants detect and learn arbitrary relations involving speech versus nonspeech intersensory relations involving the face and voice? Clearly, many questions remain to be answered concerning infant's intermodal perceptual abilities.

Matching speech sounds and articulation

There is some evidence that infants are able to match equivalent information in facial and vocal speech. Dodd (1979) found that 2- to 4-month-olds looked longer at a woman's face when the speech sounds and lip movements were in synchrony than when they were asynchronous. However, detection of synchrony does not reveal knowledge of the match between phonemes and articulatory movements.

Kuhl and Meltzoff (1982; 1984) conducted the first study that specifically examined infants' ability to detect a match between articulatory movements and vowel sounds. Using the preferential looking technique, Kuhl and Meltzoff presented 4.5-month-olds with filmed images of a woman articulating the vowels /i/ and /a/. During the familiarization phase, each face was presented without sound for 10 s;

during the test phase, the sound track for one of the vowels was played for 2 min. The woman's face was framed by black cloth to occlude neck, hair, and ears and the auditory and visual stimuli were aligned so that temporal synchrony was equal for both images. Infants looked significantly longer at the face that matched the sound. There were no other significant effects: no preference for the face located on the infant's left or right side, or for the /a/ or the /i/ face. These results were replicated with a new set of 4.5-month-olds and with a new pair of vowels (/i/, /u/) (Kuhl & Meltzoff, 1988).

In order to identify the stimulus features that are necessary and sufficient for detecting cross-modal equivalence, Kuhl and colleagues (Kuhl & Meltzoff, 1984; Kuhl, Williams, & Meltzoff, 1991) selected pure tones of various frequencies that isolated a single feature of a vowel without allowing it to be clearly identified. Adults and 4.5-month-old infants were presented with tasks that assessed their abilities to relate the pure tones to the articulatory movements for /i/ and /a/. Although adults could still detect the match, infants showed no preference for the match when the auditory component was reduced to simple tones. These results suggest that the gaze preferences observed with intact vowel stimuli were not based on simple temporal or amplitude commonalities between audio and visual streams, but rather were likely based on matching of more complex spectral information contained in the auditory component with articulatory information. Kuhl et al. (1991) claim that since spectral information, unlike temporal and amplitude information, depends largely on articulatory changes, sensitivity to the relationship between spectral information and visual speech is based on linking phonetic primitives. Therefore, 4-month-olds may be

sensitive to structural correspondences between the acoustic and visual properties of articulation.

Infants' ability to detect the match between mouth movements and speech sounds has been extended in three independent studies. MacKain, Studdert-Kennedy, Spieker, and Stern (1983) presented infants with two simultaneous video displays of two women articulating three pairs of disyllables: /mama lulu/, /bebe zuzi/, and /vava zuzu/. Infants between 5- and 6-months of age looked longer at the sound-specified display, but only when the matching face appeared on the right-hand side. The authors suggested that looking to the right side facilitates intermodal speech perception and thus indicates that in infancy the left hemisphere is predisposed to process cross-modal speech-speaker correspondences. However, asymmetries of lateral gaze have not been validated as an index of cerebral lateralization (Rose & Ruff, 1987).

Walton and Bower (1993) replicated Kuhl and Meltzoff's (1988) visual preference findings with the vowels /i/ and /u/ using an operant-sucking method. Four-month-olds sucked more to receive the appropriate face-voice pairings. In a second study, the /u/ face was paired with the /u/ sound (match), the /i/ sound (impossible), or a French /y/ sound (possible but unfamiliar). Infants between the ages of 6- and 8-months sucked significantly more to receive the possible face-voice pairs than to receive the impossible pair. In fact, novelty seemed to enhance the appeal of articulatory possibility. Thus familiarity did not appear to be the cause of infants' preference for matching face-voice pairings. Finally, results from infants' "first look" data suggested

that infants did not have to work out that a mismatch was impossible, but that it was perceived rapidly.

Patterson and Werker (1999) replicated and extended Kuhl and Meltzoff's (1982) vowel matching effect with 4.5-month-old infants. Infants looked longer at the sound-specified face even when presented with more complete visual displays (faces were shown from the shoulders up, revealing hair and some clothes) and with male as well as female stimuli. Similar to Kuhl and Meltzoff (1982; 1984), Patterson and Werker found no preference for the /i/ or the /a/face, no overall right/left side preference, no infant sex differences, and no preference for either the female or male stimuli. Thus more complex visual stimuli do not substantially impede 4.5-month-olds' ability to detect audio-visual matches based on phonetic information.

Indirect evidence for infants' sensitivity to a speaker's mouth movements has been obtained from studies of infants' imitation of vocalizations. Legerstee (1990) examined the role of audition and vision in eliciting early imitation of speech sounds. Only infants who were exposed to matching audio-visual information were observed to imitate the vowels. Kuhl et al. (1991) also observed that infants differentially imitated speech versus nonspeech sounds; moreover, when infants were listening to speech they were most likely to imitate the auditory signals they heard (Kuhl & Meltzoff, 1984; 1988). In Patterson and Werker (1999), infants also showed articulatory imitation in response to the matching face/voice stimuli. Evidence of early sensitivity to the audio-visual concomitants of vowels raises the possibility that infants younger than 4.5-months of age may also be able to link audio and visual information. To date, no

studies have examined infants' ability to match phonetic information in the face and voice at ages younger than 4 months.

Summary and Present Studies

Together, results from studies of infants' detection of audio-visual structural correspondences and vocal imitation provide converging evidence for the intermodal organization of speech in early infancy and have important implications for the acquisition of linguistic, social, and emotional skills. If prelinguistic infants are sensitive to the temporal and structural congruence of audio-visual speech information then a stronger argument can be made for the existence of invariant phonetic information across both facial and heard speech and/or a specialized module which facilitates the integration of audio-visual speech information (Liberman & Mattingly, 1985).

It seems likely that intermodal perception may play a critical role in the development of speech perception and production. However, very little research has examined how intermodal speech perception differs from intermodal perception of nonspeech relations in faces and voices. Of particular interest, if at 4.5 months of age invariant phonetic information is available in both face and voice, it is critical to examine if infants of the same age can match nonspeech events (e.g., gender) in the same faces and voices. If 4.5-month-olds can match gender in face and voice, this would lend support to claims that speech perception is no more special than other kinds of perception. However, if infants cannot match gender in the same faces and voices, this would suggest that, compared to phonetic matching, gender matching requires

more experience and learning of face-voice relations that are arbitrary and culture specific.

It might be the case that infants attend equally to all properties of audio-visual events. Alternatively, some properties may guide infants' learning of audio-visual associations more than others. Furthermore, there might be developmental changes in the extent to which different properties lead to learning of different bimodal relations. In other words, despite early sensitivity to amodal properties, one might observe developmental changes in the influence these amodal properties have on infant perception of intermodal relations. This developmental change may parallel infants' perception of increasingly specific intermodal relations as they grow older. To the extent that developmental change towards increasing specificity varies with amodal properties, one might observe greater changes in infants' reliance on some amodal properties than others. In the case of phonetic and gender information, evidence suggests that infants' perception of phonetic information might be more advanced than gender information at 4.5 months of age.

In this dissertation, a series of nine studies explores when and how infants match phonetic and gender information in dynamic faces and voices. In Section I, Experiment 1 examines 4.5-month-old infants' ability to match gender information in face and voice using the same stimuli from previous vowel matching studies. Experiments 2 through 4 place vowel and gender information in conflict to see if infants prefer certain kinds of information over others in the audiovisual presentation of faces and voices. Experiments 5 and 6 determine when infants reliably match gender information in the

face and voice. Section II explores whether matching phonetic information in face and voice is based on featural or configural information. Two experiments examine infants' ability to match phonetic information when the entire face is inverted (Experiment 7) or when only the mouth is inverted (Experiment 8). Finally, in Section III, Experiment 9 attempts to replicate the phonetic matching effect with 2-month-old infants.

Experiment 1: Gender matching in 4.5-month-old infants

Previous work has confirmed that 4.5-month-old infants are able to match vowel information in the face and voice (Patterson & Werker, 1999). When presented with side-by-side displays of either a female or a male articulating the vowels /i/ and /a/ in synchrony and an auditory vowel that matched one of the faces, infants aged 4.5 months looked significantly longer at the face that matched the heard vowel. The purpose of Experiment 1 was to determine whether 4.5-month-olds also match gender[4] information across dynamic displays of the same faces and voices used in the previous vowel matching study (Patterson & Werker, 1999). If 4.5-month-olds detect concomitant gender information in face and voice, they should look longer at the face that matches the gender of the heard sound. The use of isolated syllables and faces matched as closely as possible on all variables except gender may allow infants to focus on the structural features of gender in the face and voice.

[4] The term "gender" in this case is a socio-cultural construct based in part on sex and refers to sex-based differences between people.

Method

Participants

Mothers were recruited from a local maternity hospital shortly after giving birth or they responded to an advertisement in the local media. The final sample consisted of 64 infants, 32 male and 32 female, ranging in age from 19.1 to 24.7 weeks (M=21.1 weeks, SD=2.1 weeks). An additional 31 infants were tested and excluded from analyses due to fussiness (9), not looking at both stimuli during Familiarization (7), total looking time less than 1 min (7), looking at the same screen for the entire Test phase (2), equipment failure (4), and mother interference (2). Infants had no known visual or auditory abnormalities, including recent ear infections. Infants who were at-risk for developmental delay or disability (e.g., pre-term, low birth weight) were not tested.

Stimuli

The same stimuli were used as in Patterson and Werker (1999). Multi-media computer software (mTropolis, version 1.1) on a Macintosh 7300 was used to combine, control, and present digitized audio and visual stimuli. Infants were shown two filmed images displayed on separate side-by-side computer monitors of a female and a male face articulating the same vowel (either /a/ or /i/) in synchrony. The sound track corresponding to the articulated vowel was presented through a speaker (Sony SRS-A60) midway between the two images. Since infants can detect face-voice correspondences based on temporal cues, the two visual images were presented in synchrony and the sound was aligned with the images so that it was equally synchronous with the onset of both mouth movements.

Female and male faces were selected for similar coloring (Caucasian, fair hair) and attractiveness. As illustrated in Figure 1, the female had shoulder-length hair and the male's hair was all-one-length extending just below his ears. Both the female and male were filmed against a black background, both wore white turtlenecks, and neither

Insert Fig 1 about here

wore jewelry or make-up. First, the male was filmed producing the vowel /a/ to the beat of a metronome set at 1 beat per 3 s. This 2 min recording was then played back over a TelePrompter and all other vowels (male /i/ and female /a, i/) were produced in synchrony with the male's /a/.

As in Kuhl and Meltzoff (1984) and Patterson and Werker (1999), a different male and female were selected to record the audio stimuli. Different voices were used to ensure that there were no idiosyncratic cues linking a specific voice to a specific face. Audio recordings were made in a sound-proof recording booth using a studio-quality microphone and were recorded onto audio tape. Speakers were asked to articulate the vowels /i/ and /a/ with equal intensity and duration.

One visual /a/, one visual /i/, and one instance of each vowel sound for both female and male stimuli were chosen by three judges who rated what they deemed to be the five best visual and audio stimuli[5]. The facial images were chosen such that

[5] Kuhl and Meltzoff (1982) chose 20 audio and visual /a/s and /i/s to make two film loops. I digitized the audio and visual stimuli onto an I-Omega CD, thus the file size was limited such that only three instances of each audio and visual stimulus could be chosen. When transferring these files to the multi-media authoring program, one instance of each audio and visual stimulus was chosen in order to speed up running of the program.

duration of individual articulations fell within a narrow range that overlapped for the two vowels, the head did not move, and one eye blink occurred after each articulation. For the female, the length of time that the lips were parted was .94 s for /a/ and .95 s for /i/. For the male, this duration was 1.27 s for /a/ and 1.28 s for /i/. A comparable process was used to select the audio stimuli. Since duration of mouth opening can be longer than sound duration but not vice versa, we ensured that the vowel sounds were of the same or shorter duration than the mouth opening. For the female, duration of the sound was .61 s for /a/ and .63 s for /i/. For the male, this duration was .62 s for /a/ and .73 s for /i/.

The films and audio files were digitized and entered into a customized computer program (mTropolis, version 1.1) which locked the appropriate faces in phase. Next, the appropriate audio file was linked to each visual stimulus 1 s (15 frames) after the mouth first started to move. Each articulation was repeated to form a continuous series of articulations occurring once every 3 s. When displayed on the monitors the faces were approximately life-size, 17 cm long and 12 cm wide, and their centers were separated by 41 cm. The sounds were presented at an average intensity of 60 $\pm$ 5 dB SPL.

Equipment and Test Apparatus

The stimuli were presented on two 17" color monitors (Acana 17P) in the testing room. Black curtains covered the wall so that only the monitor screens and the camera lens, positioned between and above the two monitors, were visible. The infant was seated 46 cm from the facial displays, which subtended 29 degrees of visual angle, in an

infant seat secured to a table and the caregiver was seated behind the infant. The speaker was centered midway between the two monitors behind the curtain. During testing, a 60-Watt light in a lamp shade was suspended 1 m 10 cm above the infant.

Procedure

The experimental procedure involved two phases: Familiarization and Test. During the Familiarization phase, the visual stimuli were presented without sound for 27 s. First, each visual stimulus (the articulating male and female face) was presented alone, one on each monitor, for 9 s each. During the final 9 s of the Familiarization phase, both articulating faces were presented simultaneously without sound[6]. Both stimuli were then occluded for 3 s before the Test phase began. During the 2 min Test phase, both faces were presented simultaneously and one of the two sounds (either /a/ or /i/) was played repeatedly (see Figure 2). Sound presented, left-right positioning of the two faces, order of familiarization, and infant sex were counterbalanced across infants.

Insert Figure 2 about here

Scoring

Coding was performed using a Panasonic video recorder which allowed frame-by-frame analysis. Coders were undergraduate students who were blind to the stimuli presented to the infant. Inter-observer reliability was assessed by re-scoring 25% of the

[6] Kuhl and Meltzoff (1982) did not include the simultaneous presentation of both faces in their Familiarization phase. This phase is typically included in studies of word comprehension (e.g., Hirsch-

participants. Duration of gaze was scored for each second when the infant was looking either at the right or at the left monitor. Individual gaze-on seconds were summed for each display and divided by the total time spent looking at the displays to obtain the percentage of total looking time (PTLT) spent on each display during the Test phase as well as for the 9 s period of the Familiarization phase where both faces were presented simultaneously. PTLT to the match was also calculated for the first and second minutes of the Test phase separately. Finally, the longest look to the match and the mismatch was recorded for each infant and summed across infants. The percentage agreement for each second in the sampled periods ranged from 95.9% to 99.4% (M=97.9%) for infant looking.

Results and Discussion

According to a paired t test, no side bias was evident during the Familiarization phase (p=.40). Infants spent 74.2% of the total Test phase looking at either of the two faces. Overall, infants aged 4.5 months showed no evidence of matching on the basis of gender. As illustrated in the first bar of Figure 3, infants spent, on average, 52.4% of the total looking time at the match, which was not significantly greater than chance (50%; p=.51). Of the 64 infants tested, 40 looked longer at the match than the mismatch (binomial test, p=.06). Using percentage of total looking time (PTLT) to the match as the dependent variable, an omnibus ANOVA was conducted to explore any possible interactions among four primary variables (Sex of infant, Side of match, Heard vowel, and Gender of face/voice). All main effects and interactions were nonsignificant.

Pasek & Golinkoff, 1992). The logic behind including this phase is to teach infants that both displays can be on simultaneously and it can be used as a check for infant side bias.

Other factors that were counterbalanced in the design were entered into separate one-way ANOVAs; there were no significant differences in looking time to the match based on order of familiarization or side of gender.

Insert Figure 3 about here

When infant looking was examined during the first and second minutes of the Test phase there was still no evidence of matching based on gender information [PTLT(min1)=51.8, p=.65; PTLT(min 2)=55.0, p=.21]. Similarly, infants' longest looks to the match (19.31 s) versus the mismatch (19.56 s) were very similar. Therefore, the 4.5-month-olds tested in this study, an age at which infants are able to match phonetic information in the face and voice, were not able to match gender information using the same stimuli.

CONFLICTING AUDIO-VISUAL INFORMATION

Research has shown that infants between 4 and 6 months of age are highly sensitive to and interested in vowel information (Meltzoff & Kuhl, 1994; Polka & Werker, 1994). It is possible that the 4.5-month-olds in Experiment 1 failed to match on the basis of gender because they attempted instead to match the vowel information in face and voice. That is, vowel information may have been more salient for infants than gender information. Since the heard vowel matched both faces in terms of phonetic information, infants may have stopped looking for a "match". In the next three experiments the plausibility of this hypothesis was tested by placing vowel and gender

information in conflict to see if infants tend to match on the basis of vowel over gender in the audio-visual presentation of faces and voices.

Experiment 2: Gender Matching with Discrepant Vowel

If 4.5-month-old infants can match gender information in the face and voice but were distracted in Experiment 1 by the presence of matching phonetic information, then infants may reveal preferential looking if there is no matching phonetic information. To test this possibility, infants were presented with a male and a female face, each articulating the same vowel sound (either /a/ or /i/). One of the faces matched the gender of the voice and the other did not. Of critical importance, the heard vowel did not match the articulated vowel on either of the two faces. For example, an infant might be shown displays of a male and a female face each articulating the vowel /a/. They would then hear a female voice saying /i/. By removing the possibility of a phonetic match, I hoped to more unambiguously focus the infants' attention on the gender of the face and voice.

Method

Participants

The final sample consisted of 32 infants, 16 male and 16 female, ranging from 17.5 to 23.1 weeks (M=19.4 weeks, SD=1.9 weeks). An additional 15 infants were tested but excluded from analyses due to fussiness (5), failing to look at both screens during Familiarization (4), looking at the same screen for the entire Test phase (2), equipment failure (1), and mother interference (3). Equipment, test apparatus and scoring were identical to Experiment 1.

Stimuli and Procedure

As in Experiment 1, infants were seated in front of two side-by-side visual displays, one presenting a female face and the other presenting a male face, both articulating the same vowel sound (either /a/ or /i/) in synchrony. However, unlike Experiment 1, the heard vowel matched one of the articulating faces in terms of gender but neither in terms of vowel. All other aspects of the stimuli and procedure were identical to Experiment 1. The percentage agreement for each second in the sampled periods ranged from 96.3% to 98.9% (M=98.0%) for infant looking.

Results and Discussion

Infants showed no evidence of side bias during the Familiarization phase (paired t test, p=.45). Overall, infants spent 84.2% of the Test phase fixating one of the two faces. As illustrated in the first bar of Figure 4, infants did not show a visual preference for the face that matched the gender of the voice despite non-matching

Insert Fig 4 about here

phonetic information (PTLT=56.28; p=.25). A 4-way ANOVA (Infant sex, Side of match, Heard vowel, Gender of voice) revealed no significant main effects or interactions. The effect was similar across the first minute (PTLT=56.11) and the second minute (PTLT=56.61) of the Test phase and only 19 out of 32 infants looked longer at the gender match (binomial test, p=.35). Finally, infants' longest look to the match (25.24 s) versus the mismatch (19.28 s), although suggestive, did not differ significantly according to a paired t test (p=.098).

In Experiment 2, I attempted to remove vowel as a source of matching by presenting different phonetic information in face and voice. Infants aged 4.5 months still showed no preference for the gender match. Both PTLT and longest look to the match were marginally longer than in Experiment 1, but the effect was not significant. This finding suggests that, even when a potentially distracting phonetic match is not possible, infants still do not match gender cues in face and voice. In reminder, earlier work (Patterson & Werker, 1999) has shown that 4.5-month-old infants were able to match vowel cues in the same faces and voices when the gender cues were congruent. The results of Experiments 1 and 2, in comparison to the earlier work, show that infants match phonetic information before they are able to match gender cues in the faces and voices used in this dissertation. Moreover, infants' failure to match on the basis of gender even when matching vowel information is eliminated raises the possibility that 4.5-month-olds may not even detect the gender-relevant information in my stimuli. In the next two experiments, I explore this further by seeing whether 4.5-month-olds can match phonetic information in the face and voice when the gender information in the faces is incongruent with the heard voice, or whether the presence of conflicting gender information disrupts vowel matching.

Experiment 3: Vowel Matching with Discrepant Gender

The purpose of Experiment 3 was to see if infants still match faces and voices based on vowel information when the gender of the voice does not match the gender of either face. Infants aged 4.5-months match phonetic information when the face and voice are of the same gender (Patterson & Werker, 1999). Moreover, bimodal speech

perception in adults is unaffected by mismatching gender. Green and Kuhl (1991) presented adults with stimuli wherein a male voice saying /ba/ had been dubbed onto a female face articulating /ga/ and the subject perceived "da" as in the classic "McGurk" finding (McGurk & MacDonald, 1976). Conflicting gender information in this case had no significant impact on the McGurk effect in adults even though the discrepancy was easily detected by the subjects. If infants are like adults, they should be able to match phonetic information even when the gender information in face and voice is incongruent.

Method

Participants

The final sample consisted of 32 infants, 16 male and 16 female, aged 16.2 to 22.3 weeks (M=19.8 weeks, SD=1.4 weeks). An additional 12 infants were tested but excluded from analyses due to fussiness (5), not looking at both screens during familiarization (3), looking at the same screen for the entire Test phase (1), mother interference (2), and equipment failure (1). Stimuli, test apparatus and scoring were identical to Experiment 1.

Procedure

The procedure was identical to Experiment 1 except that infants were presented with two faces of the same gender and the soundtrack corresponded to the vowel articulation of one of the faces but was of the opposite gender. Therefore, infants were presented with either side-by-side male or female faces, each articulating a different vowel sound (either /a/ or /i/) in synchrony. The gender of the heard sound did not

match either of the seen faces, however, the heard vowel matched one of the two articulating faces. For example, an infant might see two female faces, one articulating the vowel /i/ and one articulating the vowel /a/, but would hear a male voice saying /a/. In this condition, if the infant can match vowel information despite conflicting gender he/she should look longer at the face articulating /a/. The percentage agreement for each second in the sampled periods ranged from 95.7% to 99.5% (M=98.2%) for infant looking.

Results and Discussion

No evidence of side bias was observed during the Familiarization phase (paired t test, p=.38). Infants spent 78.3% of the total Test phase fixating one of the two faces. As illustrated in the second bar of Figure 4, infants spent significantly more time looking at the face that matched the heard vowel even though gender of the voice did not match the gender of the seen face (PTLT=62.5, t(31)=2.39, p=.02). Similarly, the effect was present in both the first minute (PTLT=61.0; t(31)=1.99, p=.03) and second minute (PTLT=64.5; t(31)=2.91, p=.01) of the Test phase and 24 out of 32 infants looked longer at the sound-specified face (binomial test, p=.025). Infants' longest looks during the Test phase were significantly longer to the match (32.4 s) versus the mismatch (13.6 s) according to a paired t test (t(31)=2.61, p<.04).

A 4-way ANOVA (Infant sex, Side of match, Heard vowel, Gender of voice) revealed significant main effects for Side of match (F(1,16)=25.2, p<.01) and Heard vowel (F(1,16)=19.8, p<.01). Overall, infants looked longer at the match when it was on the right side (PTLT=75.0) versus the left side (PTLT=48.5) and when the heard vowel

was /a/ (PTLT=73.7) versus /i/ (PTLT=51.5). There was also a significant Side of match x Heard vowel interaction ($F(1,16)=13.2$, $p<.01$); infants looked longer at the match when it was on the left-hand side if the heard vowel was /a/ rather than /i/. A significant Side of match x Gender of voice interaction ($F(1,16)=4.75$, $p<.045$) revealed that infants looked longer at the match when it was on the left-hand side if the gender of the heard voice was female rather than male. Finally, there was a significant 3-way interaction among Infant sex x Side of match x Gender of voice ($F(1,16)=7.62$, $p<.014$); boys looked longer at the match on the right hand side when the female voice was played and longer at the match on the left hand side when the male voice was played. Girls' looking time to the match on the left and right side was not influenced by gender of heard voice.

These results, like those reported with adults (Green & Kuhl, 1991), suggest that 4.5-month-old infants are able to ignore incongruent gender information in order to match on the basis of phonetic information in face and voice. This stands in contrast to the results from Experiment 2 in which infants of this age showed no evidence of matching based on gender when there was incongruent phonetic information. Thus, Experiment 3 provides additional evidence that phonetic information is a particularly salient cue for infants at 4.5 months of age. It should be noted, however, that while the majority of infants preferred to look at the face that matched the heard sound in terms of vowel and the PTLT (62.5) was very close to that reported in Patterson and Werker (1999) for phonetic matching with congruent gender (62.7 for male stimuli and 64.7 for female stimuli), there was considerably more variability in Experiment 3 than in the

Patterson and Werker (1999) vowel matching study. In that study, there was no significant side bias or Heard vowel effect, nor were there any significant interactions. The side bias, Heard vowel effect, and interactions reported in Experiment 3 reveal that some infants may have been confused by the mismatching gender information and resorted to side preferences or other idiosyncratic behaviors. This raises the possibility that gender information may be noticed, to some extent, by infants at 4.5 months of age, even if they cannot yet use gender information to form an integrated percept.

Experiment 4: Vowel and Gender Information in Conflict

If infants are born with a unity of the senses and expect certain sounds to be concomitant with certain events (Morongiello, 1994), then conflicting audio-visual information, particularly when synchrony is maintained, might be especially distressing or attention-getting. If infants do not notice gender at all in my stimuli, it should not provide a source of conflict. However, if, as suggested by Experiment 3, gender is detected, it might interfere with vowel matching. To test this hypothesis, in Experiment 4 gender *and* vowel information were placed in full conflict to see if infants show any matching—be it gender or phonetic—under these conditions. For example, an infant might *see* a male face articulating /a/ and a female face articulating /i/ and *hear* a male voice saying /i/. Therefore, infants could look at the face that matched the heard sound in terms of vowel, or in terms of speaker gender, or neither. If 4.5-month-olds do not match gender information with my stimuli because they do not perceive gender cues as such then, when vowel and gender cues are placed in direct conflict, mismatching gender should not interfere with infants' ability to match based on vowel

information. However, as suggested by Experiment 3, if infants can at some level detect equivalent gender information in face and voice, one might expect this information to interfere with their ability to match based on vowel information. To the extent that conflicting audio-visual information is disruptive, if gender information is detected then it should interfere maximally with infants' ability to match on the basis of phonetic information in this full conflict manipulation.

Method

Participants

Recruitment procedures were identical to Experiment 1. The final sample consisted of 32 infants, 16 male and 16 female, ranging in age from 17.9 to 22.2 weeks (M=19.3 weeks, SD=1.3 weeks). An additional 13 infants were tested but excluded from analyses due to fussiness (4), not looking at both stimuli during Familiarization (3), total looking time less than 1 min (3), and looking at one screen for the entire Test phase (3).

Stimuli and Procedure

All aspects of the stimuli, test apparatus, and procedure were identical to Experiment 1 with the following exceptions. Infants were shown side-by-side images of a male and a female face on separate monitors, each articulating a *different* vowel (either /a/ or /i/). The sound track corresponded to the gender of one of the faces but matched the other face in terms of vowel articulation. The two visual images were temporally synchronous with the vowel sounds. The percentage agreement for each second in the sampled periods ranged from 94.9% to 99.6% (M=98.3%) for infant looking.

Results and Discussion

No side bias was evident during the Familiarization phase (paired *t* test, *p*=.81). Overall, infants spent 87.5% of the total Test phase fixating one of the two faces. During the Test phase, infants spent approximately equal amounts of time fixating the gender match (PTLT=51.8) and the vowel match (PTLT=48.2) according to a paired *t* test; thus, infants did not show a visual preference for either gender or linguistic information (see Figure 4). Out of 32 infants, 16 looked longer at the gender match while 16 looked longer at the vowel match.

All statistics were conducted using PTLT to the vowel match as the dependent variable. A 4-way ANOVA (Side of vowel match, Infant sex, Gender of voice, Heard vowel) revealed no significant main effects for Infant sex, Gender of voice, or Heard vowel; however, there was a significant main effect for Side of vowel match ($F(1,16)=7.37$, $p=.015$) and its interaction with two other variables, Infant sex ($F(1,16)=6.57$, $p=.02$) and Heard vowel ($F(1,16)=5.98$, $p=.026$). Boys, but not girls, looked longer at the vowel match when it was on the right-hand side rather than on the left-hand side and, overall, infants looked longer at the vowel match when it was on the right-hand side if the heard vowel was /i/ rather than /a/. Paired *t* tests revealed that PTLT spent on the vowel match did not significantly differ from the first minute (45.0) to the second minute (53.7, $p=.53$) nor did infants' longest look to the vowel match (21.8 s) differ from longest look to the gender match (21.2 s, $p=.83$).

It seems that simultaneously conflicting vowel and gender information disrupted infants' ability to match on the basis of phonetic information. This finding is consistent

with the hypothesis that 4.5-month-olds do notice gender-relevant cues but are unable to use the auditory specification of gender to guide visual exploration of facial cues when both gender and phonetic information are varying in the face. Several researchers (e.g., Bahrick & Pickens, 1994) have described a lag between infants' noticing information and their ability to use auditory information to guide visual exploration of that information when irrelevant visual information is simultaneously present. The results from Experiments 3 and 4 provide support for that notion.

If the hypothesis that 4.5-month-old infants are at least noticing mismatching gender is correct, and if detection precedes more functional use, one would expect to see evidence of gender matching emerging at a slightly older age. The next experiment was designed to test this hypothesis with slightly older infants.

Experiment 5: Gender Matching in 6-month-olds

The only previous study examining infants' ability to match dynamic audio and visual information based on gender cues reported evidence of matching at 6 months of age (Walker-Andrews et al., 1991). The results reported in the current set of studies suggest that infants at 4.5 months of age detect gender information in face and voice but are not yet able to use it to guide their looking preferences. Thus, by 6 months of age, infants may be able to match on the basis of gender using my stimuli. Experiment 6 was designed to specifically test this hypothesis by testing 6-month-olds in the same task used in Experiment 1.

Method

Participants

Recruitment procedures were identical to Experiment 1. The final sample consisted of 32 babies ranging in age from 24.9 to 27.1 weeks (M=26.6 weeks, SD=1.9 weeks). An additional 13 infants were tested but excluded from analyses due to fussiness (7), not looking at both stimuli during Familiarization (4), and equipment failure (2). Equipment, procedure and scoring were identical to Experiment 1. The percentage agreement for each second in the sampled periods ranged from 96.2% to 99.5% (M=98.1%) for infant looking.

Results and Discussion

A paired t test indicated that infant looking during the Familiarization phase was not biased to either the right or the left side. Overall, 6-month-olds spent 77.4% of the total Test phase fixating one of the two faces. The ability of 6-month-old infants to match faces and voices based on gender appeared to be stronger than at 4.5 months but was still not significantly above chance and was not as robust as vowel matching at 4.5 months. As illustrated in Figure 3, 6-month-olds spent 58.1% of the total looking time on the match ($t(31)=1.66$, $p=.11$) and of the 32 infants tested 22 looked longer at the match (binomial test, $p=.25$).

A paired-sample t test indicated that looking time to the match was similar across infants during the first minute (57.0%) and second minute (58.3%) of the Test phase ($p=.51$). Overall, paired-sample t tests indicated no significant difference between the longest look to the match (16.9 s) versus the mismatch (11.8 s) during the Test phase

(p=.43); however, the duration of longest look for girls was significantly longer to the match (18.9 s) versus the mismatch (9.2 s, $t(15)$=2.16, p=.01).

A 4-way ANOVA (Infant sex, Side of match, Heard vowel, Gender of voice) revealed a significant main effect for Side of match ($F(1,16)$=8.05, p<.01); infants looked longer at the match when it was on the right side (PTLT=70.4) versus the left side (PTLT=45.7). As shown in Table 1, Humphrey and Tees' (1980) correction procedure was applied to the PTLT, however, the correction did not change the significance of the results. All other main effects and interactions were nonsignificant. Separate 1-way ANOVAs revealed no preference for either the female or male face or voice nor was there any significant effect for side of gender or order of familiarization.

Insert Table 1 about here

These results, particularly the significant effect for longest look duration in girls, suggest that 6-month-old infants may be on the verge of being able to use knowledge of gender information in my stimuli to guide looking preferences. Several studies (e.g., Fenson et al., 1994; McClure, 2000) that have examined infant speech/language development and face perception in general have found that girls are often precocious in their perceptual abilities relative to boys. The findings from Experiment 5 support this claim. The final study was designed to determine when infants show evidence of matching gender information in dynamic faces and voices.

Experiment 6: Gender Matching in 8-month-olds

It may be that 4.5- and 6-month-old infants are unable to match gender with my brief auditory stimuli and when typical cultural cues to visual gender are minimized. In reminder, the 6-month-olds who were reported to match face and voice based on gender in Walker-Andrews et al. heard 20 s clips of a nursery rhyme whereas infants in my studies heard an isolated vowel. By 8 months of age, infants have had more experience with different-gender faces and voices and, by 7- to 10-months of age, can categorize faces and voices based on gender. Thus, a stronger preference for the gender match was expected at 8-months of age.

Method

Participants

Recruitment procedures were identical to Experiment 1. The final sample consisted of 32 infants, ranging in age from 33.6 to 38.3 weeks (M=36.7 weeks, SD=2.1 weeks). An additional 19 infants were excluded from analyses due to fussiness (8), not looking at both stimuli during Familiarization (4), total looking time less than 1 min (5), and looking at the same screen for the entire Test phase (2). All aspects of the stimuli, equipment, procedure and scoring were identical to Experiment 1. The percentage agreement for each second in the sampled periods ranged from 94.2% to 99.4% (M=97.2%) for infant looking.

Results and Discussion

According to a paired t test, there was no significant side bias during the Familiarization phase. Overall, 8-month-olds spent 79.1% of the Test phase fixating one

of the two faces. Unlike the younger infants, 8-month-olds looked significantly longer at the face that matched the gender of the heard voice. As shown in Figure 3, overall, infants spent 63.7% of the total looking time on the sound-specified face ($t(31)=3.44$, $p<.01$). Of the 32 infants tested, 26 looked longer at the match than at the mismatch (binomial test, $p<.01$). The effect was also present in both the first minute (PTLT=62.5; $t(31)=3.75$, $p<.01$) and second minute (PTLT=67.2; $t(31)=4.49$, $p<.01$) of the Test phase and infants' longest looks during the Test phase were significantly longer to the match (17.7 s) versus the mismatch (8.1 s) according to a paired t test ($t(31)=2.84$, $p<.01$).

A 4-way ANOVA (Side of match, Infant sex, Heard vowel, Gender of voice) revealed a significant main effect for Side of match ($F(1,16)=8.35$, $p<.01$); both sexes looked longer at the match when it was on the right hand side (PTLT=74.5) versus the left hand side (PTLT=53.5; please see Table 1). All other main effects and interactions were nonsignificant. Separate 1-way ANOVAs revealed no significant differences in looking time based on order of familiarization, side of gender and gender of voice or face ($p>.05$).

Comparing 4.5-, 6-, and 8-month-olds' ability to match gender in face and voice.

These results suggest that by 8 months of age infants show robust performance on a gender matching task. I conducted one further set of analyses to compare the PTLT to the gender match at the three ages tested in Experiments 1, 5, and 6. The mean PTLT to the match across the three age groups is summarized in Figure 3. A t test comparing the combined PTLT to the match across all three age groups to chance (50%) revealed that, overall, infants did look significantly longer at the gender match

($t(127)=2.97$, $p<.01$). A priori comparisons of PTLTs among the three age groups tested in the gender matching task were carried out by performing three independent-sample t tests. The PTLTs for the 4.5- versus 6- month-olds did not differ significantly ($t(94)=.776$, $p=.69$) nor did the PTLTs for the 6- versus 8-month-olds ($t(62)=.924$, $p=.15$). However, the PTLT for the 8-month-olds was significantly different than that for the 4.5-month-olds ($t(94)=1.815$, $p=.04$). A trend analysis revealed a marginally significant linear trend in the PTLT across Experiments 1, 5, and 6 ($F(1,125)=3.04$, $p=.06$), indicating that performance at 6 months of age was intermediate between that at 4.5 and 8 months of age. These findings for gender matching stand in contrast to the robust findings for matching of phonetic information in the same faces and voices at 4.5 months of age. These findings provide additional evidence that the ability to match phonetic information in the lips and voice develops earlier than does gender matching and may be based on the detection of amodal invariants. The next section will explore some of the possible bases for the phonetic matching effect.

SECTION II

INVESTIGATING THE BASIS OF AUDIOVISUAL MATCHING IN INFANTS

Although there has been a great deal of research specifying what kinds of audio-visual information infants can and cannot relate, there have been few inquiries into the basis of audio-visual matching. The inverted-face paradigm is frequently used to examine the basis of face recognition and processing (e.g., Yin, 1969; Farah et al., 1995) and, less commonly, to explore the basis of audio-visual speech perception (Jordan & Bevan, 1997). Changes in facial orientation may affect the perception of visual speech

because a shift in the orientation of the talker's face creates a concomitant shift in the spatial relationships among visible articulators. Recent research with adults and 6- to 10-year-olds (Schwarzer, 2000) suggests that upright faces are processed more holistically than are inverted faces. If mental prototypes for visual speech are developed through experience with upright faces and if access to these prototypes is normally achieved by encoding the position and movement of the visible articulators in upright faces, then access may be more difficult when speech is viewed in non-upright faces.

Studies using the Inverted-Face Paradigm with Adults

The face recognition literature has shown that the perception of static facial information is sensitive to facial orientation (e.g., recognition of famous faces; Rock, 1974); however, the perception of visual speech may not be affected in the same way. Massaro and Cohen (1996) examined whether an inverted view of the face influences bimodal speech perception or simply influences the information available in visible speech. Adults identified auditory, visible, and bimodal syllables (e.g., /ba/, /va/, /da/) while a computer animation of a realistic face was presented in an upright or inverted orientation. Inverting the face influenced the amount of visible information perceived but did not change the nature of information processing in bimodal speech perception. This finding supports previous findings (Campbell, 1994; Green, 1994) that visual speech that was either congruent or incongruent with auditory speech affected the perception of auditory syllables more when faces were upright than when faces were inverted. However, in contrast to the consistent and substantial effects of facial

inversion observed with static facial images, inverting dynamic talking faces appears to produce variable and sometimes minor changes in performance (Green, 1994).

Similarly, Rosenblum et al. (2000) investigated whether image manipulations known to disrupt face perception also disrupt visual speech perception. Visual and audiovisual syllable identification tasks were presented with upright faces, inverted faces, upright faces with inverted mouths, inverted faces with upright mouths, and with isolated upright or inverted mouths. Results revealed that for some visual syllables (e.g., /va/) only the upright face–inverted mouth disrupted audiovisual identification whereas other syllables (e.g., /ba/) were not disrupted in any of the face or mouth-only conditions. The authors concluded that if holistic information is used for visual speech, its use is dependent on the visual segment and may be more important for the extraction of "more extreme visual speech movements" (p.817).

Jordan and Bevan (1997) reported three experiments with adults in which the effects of facial orientation on visual speech processing were examined using a dynamic talking face presented at eight different orientations through 360 degrees. Facial orientation did not affect the identification of visual speech per se or the accuracy of auditory speech report with congruent AV stimuli, particularly when the mouth movements presented in upright and inverted faces had similar appearances. However, facial orientation did affect the accuracy of auditory speech reported with incongruent AV stimuli, particularly when visual speech presented in upright and inverted faces had a dissimilar appearance. The finding that identification and influence of visual speech in upright faces was reduced by facial inversion only in

certain conditions (i.e., when inversion altered the visual appearance of mouth movements in incongruent speech) may suggest that relational features are less important for processing visual speech than for recognizing faces.

Studies using the Inverted-Face Paradigm with Infants

No research has used the inverted-face paradigm to examine infant speech perception, however, inverted faces have been used to examine face perception in general. Unfortunately, it remains unclear whether the early development of visual processing is better characterized in terms of holistic or analytic processing. On one hand, there are studies that show that even 3- and 4-month-old infants are able to process and remember analytic as well as holistic information from visual stimuli (Ghim & Eimas, 1988). On the other hand, several studies have found evidence for the dominance of holistic processing in 2- to 3-year-olds with analytic processing only coming on-line at 4- to 5-years of age (e.g., Smith, 1989). Fagan (1972) reported that infants between 5 and 6 months of age could not differentiate between male and female faces when stimuli were rotated 180 degrees. This suggests that infants' ability to discriminate faces based on gender cues might be based on configural cues and/or familiarity with facial features as seen in upright faces. The conflicting findings are likely the result of different methods and stimuli (e.g., faces versus objects).

Based on the face recognition research, there is some reason to believe that children, and perhaps infants, may be more able to focus on specific features of the face when the face is inverted. Schwarzer (2000) found that in the context of a category-learning task there is a developmental trend from analytical to holistic face processing

between age 7 years and adulthood. At age 7 years, there was no effect of inversion on the modes of processing; however, an effect was found for 10-year-olds albeit to a weak degree. By contrast, adults showed a marked effect of inversion in categorizing faces in that they mainly processed upright faces holistically and inverted faces analytically. These results show that young children, in contrast to older children and adults, are relatively unperturbed by the inversion of a face. Similarly, Cohen and Cashon (submitted) found that 7-month-olds process facial configuration with upright faces but process independent features when the face is inverted. Thus, it may be easier for infants to identify component parts of faces if the face is inverted than if it is upright.

One of the early applications of the inverted face paradigm in the intersensory perception literature was to infants' matching of affective expression in the face and voice. Using the preferential looking procedure, Walker-Andrews (1982) presented 5- and 7-month-olds with two side-by-side films of a woman speaking in a happy manner in one film versus a sad manner in the other film. At both ages infants looked preferentially to the film that matched the soundtrack. Because lip-voice synchrony and affective information are typically confounded, Walker-Andrews further investigated the independent contribution of each to infants' ability to match filmed facial and vocal expressions. Seven-month-olds who were presented with inverted images of happy and angry faces along with a single synchronized soundtrack did not match the faces and voices, whereas those presented with upright faces did. Because synchrony information is preserved and affective information is disrupted by turning faces upside down, these results suggest that infants' matching was not based predominantly (or

only) on synchrony information. Infants apparently detected expressive information common to movements of the face and the sound of the voice. When synchrony information was minimized by occluding the mouth area, 7-month-olds continued to show significant matching whereas 5-month-olds did not (for a review see Walker-Andrews, 1997).

Kestenbaum and Nelson (1990) also examined infants' ability to match affective information in the face and voice using inverted faces. Using the infant-controlled habituation paradigm, 7-month-olds were presented with black-and-white slides of women's faces displaying various emotional expressions. Infants recognized the similarity of happy faces over changing identities and discriminated happy expressions from fear and anger when stimuli were presented upright, but not when they were inverted. When categorization was eliminated by having only one model, infants were able to discriminate happy versus fear and anger regardless of orientation. Infants were also able to discriminate toothy happiness posed by several models from non-toothy happiness and non-toothy anger when faces were upright and inverted. Thus, it seems that, in this case, differential responses were based on attending to specific features of high salience rather than to affectively relevant configurations of features.

Recently, Bahrick (1998) used the preferential looking paradigm with upright and inverted faces to examine the ability of 4- and 7-month-old infants to match dynamic faces and voices on the basis of age and maturity. Infants at both ages were capable of matching the faces and voices of adults and children when the faces were upright but not when the faces were inverted. Thus, Bahrick concluded that matching

was most likely based on perceiving configurational information in the face and/or the relative movement of features in relation to the vocal information.

In summary, it appears that inverting the face sometimes does disrupt audio-visual matching in the face and voice. Such disruptions seem to occur in situations where overall configural information is required or used for matching the face and voice (e.g., gender, affect, age). If phonetic matching in face and voice is based on featural information and if featural information in the face is not disrupted by inversion, then matching may be possible despite facial inversion. Thus, the inverted face paradigm (or simply inverting the mouth in the upright face) may be a useful method for investigating whether bimodal matching is based on featural (i.e., lip/mouth) or configural (i.e., whole face) information.

Experiment 7: Vowel Matching with Inverted Faces

Upright and inverted faces do not differ in terms of brightness, contrast, complexity, and featural information[7]; however, inversion does disrupt configural relationships in the face (Farah et al., 1995). Thus, differences in discriminating upright versus inverted faces may lie in the overall configuration of features rather than in individual features. If mental prototypes for visual speech are developed through experience with upright faces and if access to these prototypes is normally achieved by encoding the position and movement of the visible articulators in upright faces in relation to other facial features (configural information), then access may be more difficult when speech is viewed in non-upright faces. On the other hand, if vowel

[7] Featural information in the case of face processing refers to parts of a face that can be identified on the basis of natural discontinuities (Farah et al., 1995).

matching in young infants is achieved by matching phonetic information solely in the lips with the voice (featural information), it may not be significantly disrupted by an inverted face.

Inverting the face substantially alters the relative positions of the upper and lower lips (they are completely reversed). However, this alteration may not substantially alter the critical information required for matching the vowels /a/ and /i/. Jordan and Bevan (1997) reported that basic visual cues that are broadly symmetrical about the horizontal axis (e.g., open mouth, spread lips) may be encoded equally well in both upright and inverted faces, whereas resilience to facial inversion disappears when facial symmetry is absent. If this is the case for infants as well as adults, I would expect infants to be able to match /a/ (open mouth) and /i/ (spread lips) in the lips and voice despite the face being inverted. If infants in my phonetic matching task are matching based on featural information in the face and voice, then I might expect infants to look longer at the match even when the faces are inverted. However, if configural information from the entire face is needed for matching, I would expect infants to not show preferential looking.

Method

Participants

Infants were recruited in the same manner as described in previous studies. The final sample consisted of 32 infants (16 girls, 16 boys) ranging in age from 16.1 to 19.8 weeks (M=18.2 weeks, SD=1.3 weeks). An additional 7 infants were excluded from

analyses due to fussiness (2), not looking at both stimuli during Familiarization (2), total looking time less than 1 min (2), and looking at one screen for the entire Test phase (1).

Stimuli and Procedure

The same stimuli were used as in Patterson and Werker (1999) and as described in Experiment 1 with the following exceptions. A particular infant viewed two faces of the same person (female or male) and the heard voice matched the vowel articulated by one of the two faces in the gender-appropriate voice. In addition, the faces were inverted 180 degrees. Half of the infants observed a male face and half observed a female face, with the number of male and female infants per group counterbalanced. Also, the side of the vowel match was fully counterbalanced across all conditions. All other aspects of the procedure, stimuli, and scoring were identical to Experiment 1. The percentage agreement for each second in the sampled periods ranged from 95.9% to 99.4% ($\underline{M}$=97.9%) for infant looking.

Results and Discussion

A paired $\underline{t}$ test indicated that infant looking during the Familiarization phase was not biased to either the right or the left side. Overall, infants spent 80.8% of the total Test phase fixating one of the two faces. On average, infants did not look significantly longer at the face that matched the heard sound; infants spent 58.25% of the total looking time on the match ($\underline{t}(31)=1.73$, $\underline{p}=.09$) and of the 32 infants tested 20 looked longer at the match (binomial test, $\underline{p}=.08$). However, when the first and second minutes of the Test phase were examined separately, a one-sample $\underline{t}$ test indicated that infants did fixate the matching face for a significantly longer duration in the second minute

(PTLT=61.1, $t(31)=2.17$, $p=.03$) but not in the first minute (PTLT=54.5, $p=.29$). Overall, paired-sample t tests indicated no significant difference between the longest look to the match (14.9 s) versus the mismatch (10.1 s) during the Test phase ($p=.20$); however, the duration of longest look for girls was significantly longer to the match (14.5 s) versus the mismatch (7.7 s, $t(15)=2.20$, $p=.04$). A 4-way ANOVA (Infant sex, Side of match, Heard vowel, Gender of voice) revealed no significant main effects or interactions. Separate 1-way ANOVAs revealed no preference for either the female or male face or voice nor was there any significant effect for side of gender or order of familiarization.

These results suggest that 4.5-month-old infants are able to pick up some phonetic information from inverted faces and show suggestive evidence of 4.5-month-olds matching phonetic information with inverted faces, however, the inverted face may be too unfamiliar to enable robust phonetic matching as seen with upright faces.

Experiment 8: Vowel Matching with Inverted Mouths

Research that has examined adults' recognition of upright and inverted faces suggests that faces might be processed qualitatively differently in different orientations (Yin, 1969). Moreover, the accuracy and efficiency of encoding is negatively affected when faces are inverted (Bruce et al., 1991). It is possible that infants in Experiment 7 did not effectively process the inverted faces as faces and/or were not able to fully examine individual features within the whole and thus did not attend to phonetic information in the lips when the sound was played.

In Experiment 8, the face was upright however the mouth was inverted. Inverting the mouth while leaving the rest of the face upright may provide a more

familiar context to examine the role of featural and configural information in infants' ability to match phonetic information in face and voice. On the other hand, in the adult literature (Rosenblum et al., 2000), this condition is the one in which AV speech perception is most likely to be disrupted.

Research has shown that changing a single feature in a face modifies the interaction among the components and thus the particular configuration (Sergent, 1984). If infants in the phonetic matching task are matching based on featural information in the face (i.e., the mouth), then infants may also look longer at the match when only the mouth is inverted. However, if configural information from the entire face facilitates or is needed for phonetic matching, infants should not show a preference for the matching face when the mouth is inverted.

Method

Participants

Infants were recruited in the same manner as described in previous studies. The final sample consisted of 32 infants (16 girls, 16 boys) ranging in age from 19.28 to 23.14 weeks (M=20.98 weeks, SD=1.23 weeks). An additional 5 infants were excluded from analyses due to not looking at both stimuli during Familiarization (2), total looking time less than 1 min (1), and looking at the same screen for the entire Test phase (1), and mother interference (1).

Apparatus and Procedure

The same stimuli were used as in Experiment 1. Since all previous studies indicated no difference between responding to female versus male stimuli, only the

female face and voice was used in the present experiment. Computer software (Final Cut Pro, version 1.2) was used to invert the mouths within the context of the upright face. The edges of the mouth were feathered in order to blend the mouth as much as possible into the face. There was some disruption of natural shading, however, it was equivalent across both faces. As in previous studies, a multimedial computer program (mTropolis, version 1.1) was used to combine, control, and present the stimuli. All other aspects of the stimuli, equipment, and apparatus were identical to Experiment 1.

The experimental procedure was identical to that described in Experiment 7 except that a gray expanding and contracting circle was presented instead of 3 sec of black screen between the Familiarization and Test phases. The percentage agreement for each second in the sampled periods ranged from 96.1% to 99.2% (M=97.3%) for infant looking.

Results and Discussion

A paired t test indicated that infant looking during the Familiarization phase was not biased to either the right or the left side. Overall, infants spent 95% of the total Test phase fixating one of the two faces. On average, infants did not look longer at the face that matched the heard sound. Infants spent 36.8% of the total looking time on the match; thus, infants spent significantly longer fixating the mismatch ($t(31)=-2.78$, $p<.01$). Of the 32 infants tested, 22 looked longer at the mismatch (binomial test, $p=.052$). When the first and second minutes of the Test phase were examined separately, one-sample t tests indicated that infants did fixate the mismatching face for a significantly longer duration in the second minute (PTLT=35.4, $t(31)=-2.88$, $p<.01$) but not in the first minute

(PTLT=39.8, p=.07). Also, a paired-sample t test indicated a significant difference between the longest look to the match (14.3 s) versus the mismatch (29.2 s; $t(31)$=-2.37, p=.02). A 3-way ANOVA (Infant sex, Side of match, Heard vowel) revealed no significant main effects or interactions and a separate 1-way ANOVA revealed no significant effect for order of familiarization (i.e., left vs. right screen first; p=.73).

These results suggest that 4.5-month-old infants are able to distinguish inverted mouths that are articulating different vowel sounds within the context of an upright face. Since configural information in the face was disrupted in the present study, infant looking preferences may have been determined by salient featural information in the face (ie., the mouth). When adults have been presented with inverted mouths within upright faces, face processing (Valentine & Bruce, 1985) and, in some cases, audiovisual speech perception (Rosenblum et al., 2000) have been significantly disrupted. Why would inverted mouths within upright faces disrupt AV speech perception in adults but lead to a preference for the mismatch in infants?

It is possible that infants in this experiment did not process the face as a whole and focused primarily on the articulating mouth. By focusing on phonetic information in the mouth, infants may have rapidly figured out the match and become bored with it, thus leading them to seek the more novel stimulus (see Hunter & Ames', 1988 discussion of the familiarity-novelty hypothesis). Although this explanation is speculative, there is some evidence for such an explanation in the present study. For example, when the first 30 s of the Test phase was examined, infants spent 48.0% of the total looking time fixating the match, which is almost significantly greater ($t(31)$=1.97,

$\underline{p}$=.058) than the overall PTLT to the match (36.8%). Thus, it appears that infants started the Test phase looking equally at the match and the mismatch. After "figuring out" the task, they may have become bored and sought out the more novel stimulus (the mismatch). Furthermore, overall looking to either face was very high (95% of total test phase) compared to phonetic matching with upright faces (79%; Patterson & Werker, 1999), indicating that infants were interested in the inverted mouth perhaps because it was so novel.

SECTION III

Experiment 9: Matching Phonetic Information at Two Months of Age

Although several studies have shown that the ability to match phonetic information in the face and voice is robust at 4.5 months of age, no studies have examined bimodal phonetic matching in infants younger than 4 months. Soon after birth, infants are able to discriminate and categorize vowel sounds. Trehub (1973) reported that infants between 1- and 4-months of age can discriminate changes between /a/ and /i/ and between /u/ and /i/ when these vowels are isolated or follow a common consonant. Marean et al. (1992) successfully trained 2-month-olds to respond when the vowel changed from /a/ to /i/ and to refrain from responding when the vowel did not change category despite variation in the spectral cues associated with pitch and talker. The first time talker changes occurred in the absence of a vowel change, 80% of infants did not respond. Therefore, infants can discriminate vowel categories that are quite similar acoustically at an early age.

Research over the past 20 years has provided evidence for the perceptual constancy of these vowel categories in infants. Infants from 1- to 4-months of age can detect a change in the identity of synthesized vowels despite distracting variation in pitch (Kuhl & Miller, 1982) and 6-month-olds continue to detect phonetic contrasts despite changes in speaking voices from both genders that range from child to adult (Kuhl, 1979; 1983)[8]. These findings suggest that infants' speech categories are formed along similar phonetic dimensions as in adults.

Although it is clear that young infants can discriminate features within different modalities, when and how knowledge of the intermodal nature of speech is acquired by infants is not clear. There is some evidence that infants younger than 4 months of age are able to match equivalent information in facial and vocal speech. Dodd (1979) found that 2- to 4-month-olds looked longer at a woman's face when the speech sounds and lip movements were in synchrony than when they were asynchronous. However, detection of synchrony alone does not reveal knowledge of the match between phonemes and articulatory movements. Indirect evidence for infants' sensitivity to a speaker's mouth movements has been obtained from studies of infants' imitation of vocalizations. Two-month-old infants are capable of pre-speech mouth movements and imitate other mouth movements such as tongue protrusion, lip pursing, and mouth opening (Meltzoff & Moore, 1983). Despite evidence that 2-month-olds can imitate

[8] More specifically, Kuhl (1991) argues that, by 6-months of age, vowel categories may be organized around prototypical instances from the native language. Some exemplars of speech categories seem to be more "potent" than others and generalization around these "good" exemplars is significantly broader than that around "poor" exemplars. Kuhl (1991) has proposed a "perceptual magnet effect" for prototypical vowels which serves to shorten the perceptual distances between the centre and edges of the vowel category.

mouth movements, there is no research to date that has examined 2-month-olds' ability to match phonetic information in lips and voice.

Infants generally have had more exposure to female faces and voices compared to male faces and voices. If the phonetic matching effect is based on an arbitrary but natural relationship that is learned, one might expect a weaker effect with the male stimuli than with the female stimuli. In the previous vowel matching study, this was not the case with 4.5-month-olds (Patterson & Werker, 1999). These findings suggest that infants younger than 4.5 months may be able to match vowel information in the lips and voice. The fact that 4.5-month-olds were able to match phonetic information in face and voice equally well with both female and male stimuli (Kuhl & Meltzoff, 1982; Patterson & Werker, 1999) lends support to the claim that speech is especially salient for young infants and that infants may be born with biases that either allow detection without learning or that facilitate the learning of seen and heard speech.

If 2-month-olds are sensitive to the temporal and structural congruence of audio-visual speech information then a stronger argument can be made for the existence of invariant phonetic information across both facial and heard speech and/or a specialized module which facilitates the integration of audio-visual speech information (Liberman & Mattingly, 1985). However, if 2-month-olds do not detect a match between phonetic information in the lips and voice, this may be due to perceptual and/or experiential limitations. First, infants' perceptual (especially visual) abilities may not be mature enough to enable 2-month-olds to attend to the critical information. (See Appendix B for a discussion of perceptual abilities in 2-month-old infants.) Second, infants may

require more experience with articulating faces and voices before they are able to use amodal cues to guide looking preferences. Gibsonians argue that the ability to detect amodal relations should be present at birth but may be masked by the level of maturity of the infant's sense modalities, the task at hand, and the type of relationship available to perception (e.g., Walker-Andrews, 1994). Third, matching of faces and voices may involve natural and typical, rather than amodal, relations and this may require a period of learning.

Method

Participants

Mothers were recruited in the same manner described in Experiment 1. The final sample consisted of 32 infants, 16 male and 16 female, ranging in age from 7.8 to 11.1 weeks (M= 9.2 weeks, SD=1.3 weeks). An additional 19 infants were excluded from analyses due to crying (7), falling asleep (4), not looking at both stimuli during Familiarization (2), total looking time less than 1 min (4), looking at the same screen for the entire Test phase (1), and equipment failure (1). Infants had no known visual or auditory abnormalities, including recent ear infections, nor were infants at-risk for developmental delay or disability (e.g., pre-term, low birth weight). All aspects of the stimuli and procedure were identical to Experiment 7 except that the faces were right side up. The percentage agreement for each second in the sampled periods ranged from 95.1% to 99.7% ($\underline{M}$=98.1%) for infant looking.

Results and Discussion

A paired *t* test indicated that infant looking during the Familiarization phase was not biased to either the right or the left side. Overall, infants spent 79.3% of the Test phase looking at either of the two faces. Infants looked longer at a particular face when the appropriate vowel sound was heard. As shown in Figure 5, on average, infants spent 64.0% of the total looking time fixating the matching face, which was significantly greater than chance (50%), $t(31)=2.92$, $p<.01$. Of the 32 infants tested, 23 looked significantly longer to the sound-specified display than to the incongruent display (binomial test, $p<.05$). When the first and second minutes of the Test phase were analyzed separately, the effect was not present in the first minute (PTLT=60.73, $p=.075$), however, the effect was present in the second minute (PTLT=64.12, $t(31)=2.84$, $p<.01$). According to a paired-sample *t* test, infants' longest looks during the Test phase were significantly longer to the match (34.0 s) versus the mismatch (18.7 s, $t(31)=2.21$, $p=.03$). A 4-way ANOVA (Infant sex, Side of match, Speaker gender, Heard vowel) revealed no significant main effects or interactions ($p>.05$).

Insert Fig. 5 about here

When given a choice between two identical faces, each articulating a different vowel sound in synchrony, infants at 2 months of age looked longer at both a female and a male face that corresponded with the heard vowel sound. As in Kuhl and Meltzoff (1982) and Patterson and Werker (1999), the present study revealed no

preference for the /i/ or the /a/ face, no overall right or left side preference, and no infant sex differences.

As illustrated in Figure 5, the mean looking time to the match observed with infants aged 2 months (PTLT=64.0) was very similar to that observed with 4.5-month-olds using the same stimuli (PTLT=63.7; Patterson & Werker, 1999). The fact that infants are equally able to match phonetic information in face and voice at 2-months of age as at 4.5 months of age supports the hypothesis that phonetic information acts as an amodal cue available in both face and voice. The ability to match seen and heard speech at such a young age also raises the possibility of a specialized module that facilitates learning of seen and heard speech.

In general, young infants tend to have more exposure to female faces and voices compared to male faces and voices; therefore, if the matching effect is based on an arbitrary but natural relationship that is learned, one might expect a weaker effect with the male stimuli than with the female stimuli. Such a difference was not observed at 4.5 months of age (Patterson & Werker, 1999), however, one might expect to see such a difference with 2-month-olds. In the present study, no significant difference was observed in infant looking times to the female versus male stimuli. This suggests that differential exposure to male and female faces and voices does not influence infants' ability to match phonetic information in face and voice at 2 months of age. The fact that infants show evidence of matching with both female and male stimuli provides further support for the hypothesis that phonetic information is represented amodally and that

such representations require very little experience with talking faces (see Kuhl & Meltzoff, 1984).

GENERAL DISCUSSION

Previous research has shown that infants as young as 4.5-months of age can match acoustically presented vowel sounds with the appropriate facial articulation (Kuhl & Meltzoff, 1982; 1984; Patterson & Werker, 1999). These findings support claims that young infants can perceive structural correspondences between audio and visual aspects of speech input and are consistent with the possibility that phonetic information is represented amodally. It remains unclear, however, whether the intermodal perception of other biologically significant face-voice events shows the same early emergence. I chose to compare infants' ability to match phonetic versus gender information in the face and voice because gender, like phonetic information, is central to human social functioning but involves a different set of critical features than phonetic information. Thus, the current set of experiments was designed to explore the bimodal matching effect using both phonetic and gender information in dynamic faces and voices. The results from these nine experiments will be summarized and then interpreted in light of major theories of perceptual development.

Gender Matching with Dynamic Faces and Voices

In Experiment 1, 4.5-month-olds showed no evidence of matching gender information in the face and voice. I reasoned that 4.5-month-olds may be able to match gender information in face and voice but may not be interested enough in this relation to exhibit preferential looking. More specifically, although the heard vowel matched

only one face in terms of gender, it matched both faces in terms of vowel. Since infants are very interested in vowels at this age, they may have focused primarily on vowel information and may have neglected to look for a gender match. This possibility was explicitly tested in Experiment 2 by neutralizing the informational value of phonetic cues in the lips, however, infants still did not look preferentially at the gender match. Thus, it seems that 4.5-month-olds' inability to match based on gender was not entirely due to matching phonetic information in the face and voice instead.

In Experiment 3, when one of the articulating faces clearly matched the heard vowel but the gender of the voice did not match either face, infants did look longer at the face that matched the heard vowel. This shows that 4.5-month-old infants are able to neutralize conflicting gender information and recognize the vowel match. However, the unexpected interactions raised the possibility of some interference from the mismatching gender cues. To explore this possibility, phonetic and gender information were placed in full conflict in Experiment 4. Here, 4.5-month-olds did not show a visual preference for either the phonetic or the gender match. The disruption of vowel matching caused by a gender mismatch suggests that infants do notice some of the bimodal information specifying gender; however, infants at 4.5-months of age seem unable to use this information to guide their looking preferences in a gender matching task.

If, as indicated by Experiments 3 and 4, infants can detect but not match gender information in face and voice at 4.5 months of age, when does the ability to match emerge? In reminder, Walker-Andrews et al. (1991) reported that 6-month-olds can

match gender in the face and voice when the models were reciting nursery rhymes. Therefore, Experiment 5 was conducted to see if infants at 6 months match gender in the face and voice using my stimuli. Like the 4.5-month-olds, 6-month-old infants' looking time to the match was not significantly greater than chance. However, hints of preference were apparent at 6 months of age; girls' longest looks to the match were significantly longer than their longest looks to the mismatch.

In Experiment 6, an even an even older age group, 8-month-olds, was tested in the gender matching task. The ability to match gender information in face and voice was unambiguously evident in both girls and boys at 8 months of age. Preference for the gender-appropriate face was not only apparent in the overall PTLT, but also in the number of infants who looked longer at the match versus the mismatch, in PTLT to the match during the first and second minutes of the Test phase, and in infants' longest look to the match versus the mismatch.

These studies suggest that infants at 4.5 months of age do notice gender information in my stimuli. In addition, it seems that infants have some appreciation for the link between heard and seen gender since mismatching gender disrupts, partially (Exp. 3) or fully (Exp. 4), their ability to match on the basis of vowel. Nevertheless, infants do not demonstrate reliable gender matching with these stimuli until 8 months of age. Between 4.5 and 8 months of age, infants may gradually develop an integrated audio and visual representation of gender information.

It is of interest to compare the results reported here with those previously reported in the literature. In reminder, previous reports using static faces and voices

failed to find gender matching until 12 months of age. With dynamic faces, Walker-Andrews et al. (1991) failed to find robust gender matching at 4 months of age, however, matching was reported at 6 months of age. In the current set of studies, there was a trend for 6-month-old girls to match faces and voices based on gender, however, the effect was not robustly present until 8 months of age. It is noteworthy that the age at which gender matching was observed is quite similar to that reported by Walker-Andrews et al. despite the fact that our studies were conducted in different laboratories and with different stimuli. Nevertheless, it is also useful to consider why infants in this dissertation did not show robust evidence of gender matching until a slightly later age than did infants in Walker-Andrews et al.'s study.

One major difference between the current stimuli and those used by Walker-Andrews et al. (1991) concerned the audio stimuli. Walker-Andrews et al. recorded the actors' voices separately from the visual stimuli, however, the voices did belong to the visual stimuli. Thus, the voice may have been better synchronized to one of the faces, especially since the faces were articulating connected speech. Furthermore, there could have been an idiosyncratic relation between the matching face and voice (e.g., rising intonation paired with eyebrow raising) or a structurally specified relation between the physical features of the vocal organs (mouth and neck) and the sound of the voice. In the present studies, the use of voices that did not belong to either of the faces along with the use of isolated vowels allowed me to ensure that the faces and voices of both models were equally well-synchronized and that matching was not based on any idiosyncratic cues in the face and/or voice.

Second, although both my and Walker-Andrews et al.'s stimuli conveyed neutral facial and vocal affect, my stimuli were designed to minimize as many culture-specific cues as possible and to allow infants to focus on the structural features specifying gender in face and voice. To minimize culture-specific cues to gender, models were asked to remove all jewelry and makeup, the male's face was clean-shaven, and the models both wore white turtlenecks, which fully concealed any throat cues and removed any gender cues based on clothing. The only potential culture-specific cues were hair length (the female had long hair and the male had hair all-one-length just past his ears) and plucked eyebrows (female only). Since the visual stimuli eliminated many of the arbitrary face-voice associations related to gender, the infant may have had to use structural cues in the face (e.g., nose, jaw, cheekbones, skin texture) rather than culture-specific cues to match gender information in face and voice.

The voice stimuli were also created in a way that minimized cultural display rules for gender. As noted above, I chose to use isolated vowels whereas Walker-Andrews et al. used nursery rhymes. Nursery rhymes may contain more information about gender than does a single isolated vowel sound and, in particular, may contain display rules for gender in a particular culture.

The de-emphasis of culture-specific cues was instituted to increase reliance on physiognomic cues in order to assess whether gender matching, like phonetic matching, may be possible using amodal properties in the face and voice. I predicted that if infants can use amodal information to match gender in the face and voice, reliable evidence of matching should be seen at a younger age than reported by Walker-

Andrews et al. (1991). Rather than revealing an earlier emerging ability, however, infants demonstrated gender matching at a slightly later age than reported by Walker-Andrews et al. These findings suggest that matching of gender in the face and voice is not as immediately available to infants as is matching of phonetic and other amodal events and suggests that learning does play a significant role in the intermodal matching of gender. The slightly earlier case of matching reported by Walker-Andrews et al. may indicate that evidence of this learning is easier to see when there is more culture-specific information in face and voice.

It cannot be known with certainty whether the 8-month-old infants in this dissertation matched on the basis of physiognomic or culture-specific cues. Although the stimuli were designed to eliminate many culture-specific cues, a few remained. For instance, it is possible that infants at 8 months of age based looking preferences on hair length rather than on structural cues to gender. I think this is unlikely because the male used in this dissertation had hair that was longer than the stereotypical male hairstyle and equally as long as that of many females in Western culture. Nevertheless, in order to rule out this possibility, a condition with a male and female with similarly short hair (as in Leinbach & Fagot, 1993) would need to be presented to infants in a visual preference procedure. In addition, although I attempted to use stimuli that would highlight physiognomic cues in the face and voice, it is possible that by having the models wear a white turtleneck I may have eliminated a non-facial physiognomic cue to gender in the breadth of the neck and the protuberance of the Adam's apple. It would

also be useful in future work to add in this physiognomic cue to see if it facilitates gender matching.

Researchers who work from a Gibsonian model of perceptual development (e.g., Bahrick & Pickens, 1994; Walker-Andrews, 1997) have suggested that discrimination of modality-specific cues, along with the detection of amodal invariants, may precede and guide learning about arbitrary object-sound relations. Such early perceptual abilities may direct infants' attention to appropriate object-sound pairings and then promote sustained attention and further differentiation. Initial detection of an amodal relation (e.g., voice-lip synchrony, shared rhythm and tempo) could enable the infant to focus on a unitary event (e.g., the mother's face and voice). This, in turn, may lead to differentiation of more specific, arbitrarily paired audible and visible attributes (e.g., the sound of the voice with the unique appearance of the face). Bahrick (1992) reported a developmental progression across age where infants detected amodal relations at a younger age than arbitrary relations from the same events. The results from this dissertation suggest that infants detect the amodal information specifying the phonetic match in the face and voice at an earlier age than they do the gender information in those same displays. They are, however, able to discriminate the gender of the face and voice alone in similar displays (e.g., Fagan, 1976; Miller, 1983). Perhaps, as suggested by Bahrick and Pickens (1994), the early attentional bias to some amodal properties, paired with discrimination of gender cues, sets the stage for increased learning about gender with development and experience.

Although the results reported in this dissertation suggest that 4.5-month-old infants detect but do not match gender information in the face and voice, alternative explanations for the pattern of results must be considered. It was argued that because 4.5-month-olds in Experiment 3 matched phonetic information in the face and voice despite mismatching gender information, infants may have neutralized conflicting gender information. It should be noted that any single infant in Experiment 3 was presented with faces of only one gender (male or female) whereas infants in Experiments 1, 2, and 4 were presented with both a male and a female face. It is plausible that the presence of same-gender faces in Experiment 3 sufficiently reduced the information-processing load on infants so that a vowel-articulation match could be made. These infants would not have to look for (that is, process) different facial cues among the stimuli, physiognomic or otherwise, but rather they could focus on making a match between the heard vowel and the seen articulation.

Another possibility is that infants under 8 months of age have difficulty matching gender information in face and voice because of where young infants focus their attention on a visual display (see Lewkowicz, 1999). Object perception undergoes a major shift at about 7 months of age (Kellman & Arterberry, 1998). Before this time, visual processing is apparently based on an edge-insensitive process; therefore, 4.5- and 6-month-olds may not be as sensitive as older infants to the edge relationships that normally specify objects and events. Instead, they may respond primarily to kinematic properties, such as lip movements, thus enabling 4.5-month-olds to match faces and voices based on phonetic information. By 7 months of age, an edge-sensitive process

takes over and provides the infant with a richer view of the world since both kinematic properties of objects and events as well as object connectedness and forms of hidden boundaries are specified. Thus, older infants can perceive objects and events in a more detailed and accurate fashion. This perceptual shift, along with the more general process of perceptual differentiation, may in part account for the age differences in infant responsiveness to bimodal gender information. It would be of interest to explore this explanation in future research.

Phonetic Matching in Dynamic Faces and Voices

The two experiments in Section II explored whether phonetic matching at 4.5 months of age is based on featural or configural information in the face. When the entire face was inverted (Experiment 7), infants showed no significant preference overall for the face that matched the heard sound, however, hints of preference were apparent. Infants looked significantly longer at the matching face during the second minute of the Test phase and girls' (but not boys') longest look to the match was significantly longer than to the mismatch. One explanation for this weak effect is that inversion might slow down the encoding of face components due to a general slowing in visual processing (Searcy & Bartlett, 1996). Once encoding is complete, however, matching should be possible as was seen in the second minute of the Test phase.

In order to minimize disruptions in processing and to increase familiarity, in Experiment 8 only the mouth was inverted and the rest of the face was left upright. Surprisingly, infants looked significantly longer at the face that did not match the heard sound. In reminder, when the faces were upright (Patterson & Werker, 1999), infants

looked longer at the face that matched the heard vowel. The only difference between these two studies was the inverted mouth. Why did infants look longer to the match in the upright face condition, marginally longer in the inverted face condition, but longer to the mismatch in the inverted mouth condition? One possibility is that inverting the mouth disrupted configural face processing (Rosenblum et al., 2000) and forced infants to focus solely on the mouth. By focusing on the critical features needed for phonetic matching, infants may have encoded the phonetic information more readily and may have become bored with the match. Hunter and Ames (1988) have proposed that in any information-processing situation infants will show an initial preference for familiarity followed by a preference for novelty when a task becomes too easy or infants get bored. Inverting only the mouth may have disrupted holistic face processing so that infants focused primarily on the mouth, leading to an overall preference for the novel mismatch. This explanation, albeit speculative, points to the need for further research employing fine-grained analysis of infant saccadic movements to determine whether infants do indeed focus more on mouth movements than on the whole face in inverted versus upright mouth conditions.

Although familiarity effects are often reported in studies of infant speech perception (e.g., Tincoff & Jusczyk, 1999; Jusczyk & Aslin, 1995), studies using similar stimuli and familiarization procedures have reported novelty effects (e.g., Echols, Crowhurst, & Childers, 1997; Aslin, Saffran, & Newport, 1998). Exactly what causes shifts between familiarity and novelty preferences is not completely understood. Hunter and Ames' (1988) familiarity-novelty hypothesis is one explanation that has

been offered. This model would predict an initial preference for the match followed by a preference for the mismatch. To explore this possibility, I examined looking time during the first 30 sec of Minute 1. If there was an initial preference for the match in Experiment 8, it occurred in different windows for different infants and was not evident overall in the first 30 sec. Thus, if there is a trend for infants to move from a familiarity to a novelty preference, it is not straightforward.

Other explanations for familiarity and novelty preferences have been proposed. For example, Saffran (2001) has found that natural, ecologically rich stimuli tend to produce familiarity preferences while more artificial stimuli tend to result in preference for novelty. Research has shown that adults find an inverted mouth in the context of an upright face to be very unnatural and disruptive to audiovisual speech perception (Rosenblum et al., 2000). Thus, it is possible that infants found the bizarre nature of the inverted mouth to be artificial. The novelty of an inverted mouth in the context of an upright face may increase the appeal of articulatory possibility. That is, infants may have realized that the mismatch was incongruent but may have been trying to "figure out" how it was done.

The notion that young infants focus their attention on individual features more so than on overall configurations has received some support in the face perception literature (e.g., Caron et al., 1985; Cohen & Cashon, submitted). Faces have both component and configural properties and thus lend themselves to different processing strategies that are not mutually exclusive and can unfold simultaneously (Sergent, 1984). Still, the results from the two experiments in Section II provide suggestive

evidence that infants may focus on individual features more so than on configural information in the face when presented with a phonetic matching task. When considered together with the findings from the gender matching task, one can speculate that, in the first few months of life, infants succeed at matching dynamic information in the face and voice only if it is clearly conveyed by salient featural information. To the extent that gender matching relies on more configural information (as opposed to physiognomic cues), it is not surprising that it is delayed in relation to phonetic matching.

In Section III, when given a choice between two identical faces, each articulating a different vowel sound in synchrony, infants aged 2 months looked longer at both a female and a male face that corresponded with the heard vowel sound (Experiment 9). This effect was found in the percentage of total looking time spent on the matching face, the longest look to the match versus the mismatch, the number of infants who looked longer at the match versus the mismatch, and in the second minute of the Test phase. To date, this ability has only been reported in infants aged 4 months and older (Kuhl & Meltzoff, 1982; Walton & Bower, 1993; Patterson & Werker, 1999). As in Kuhl and Meltzoff (1984), the current study revealed no preference for the /i/ or the /a/ face, no overall side preference, and no infant sex differences. Therefore, the phonetic matching effect appears to be just as robust at 2 months of age as it is at 4.5 months of age. The fact that the phonetic matching effect is robust at 2 months of age provides support for an integrated, multimodal representation of articulatory and acoustic phonetic information at 2 months of age. The early emergence of this ability also suggests that

phonetic information may be represented amodally and requires relatively little experience with talking faces of different genders.

It should be noted that similar mean looking times to the match at 2- and 4.5-months of age do not rule out learning as a basis of the effect. Most 2-month-olds are not producing clear vowel sounds, so they would have had little opportunity to learn from self-produced articulations. Infants of 2 months, however, have had the opportunity to watch lips while hearing speech; thus, conclusive results addressing the question of whether phonetic matching reveals rapid learning or an inborn ability await a study with newborns. Whether or not the ability to match phonetic information in the lips and voice by 4.5-month-old infants is based on amodal specification or learning, innately-guided or otherwise, the fact that infants in Experiment 9 looked longer at the face that matched the vowel sounds suggests an integrated, multi-modal representation of articulatory and acoustic information by 2 months of age.

Relating this Work to Theories of Perceptual Development

In this section, I will discuss how the results from this dissertation can be explained by the differentiation and integration views of perceptual development. Next, I will discuss my findings in light of the motor theory of speech perception; and, finally, I will discuss how my findings can inform a more complex epigenetic-systems view of development.

As mentioned earlier, research on intermodal perception has traditionally been influenced by two opposing views of perceptual development. These views focus on whether (a) intersensory development proceeds from initially separate senses that

become increasingly **integrated** through experience, eventually resulting in coordinated multimodal perception, or (b) the development of intersensory perception is a process of **differentiation** and increasing specificity. Although research over the past twenty years has provided support for both positions, the majority of findings in the domain of audio-visual matching have been interpreted as supporting the Gibsonian "invariant detection model" (Gibson, 1969). Proponents of Gibsonian theory claim that the fact that very young infants can detect amodal invariants across different sensory modalities makes it unlikely that experience with input in two or more modalities is necessary for the initial stages of intermodal development.

According to Gibsonians, at birth, the perceptual systems are sensitive to invariant patterns in environmental information and work together as an innately coordinated system that allows the infant to experience a world of perceptual unity. A key assumption of this model is that intermodal redundancies and amodal invariants are directly available in the input and provide the infant with meaningful affordances for action (Gibson, 1969). Thus, the capacity for intermodal perception does not depend on any learned coordination between sensory modalities: infants are born with capacities to perceive unitary objects by detecting amodal properties. As infants explore the environment, they become sensitive to progressively finer distinctions among the properties of events and come to distinguish among the different sense modalities.

As mentioned previously, the results of this dissertation are generally compatible with the Gibsonian invariant detection view. The fact that 2-month-old infants match

phonetic information in the face and voice suggests that infants are sensitive to the structural correspondences that specify phonetic information in the face and voice. Furthermore, the finding that the phonetic matching effect was equally robust at 2 and 4.5 months of age makes it unlikely that infants need to learn that phonetic information from visual and audio channels is equivalent. Phonetic information appears to be amodal and is perceived directly. Further experience with such amodal relations may help infants differentiate more arbitrary relations in the face and voice (Bahrick & Pickens, 1994).

The pattern of findings for infants' matching of gender in face and voice is not as compatible with the invariant detection view, but still can be explained. Gender has been described as part of a natural, complex class of multimodal relations that are typical in the environment and are partially but not uniquely specified by amodal invariants (Walker-Andrews, 1994). For example, infants may learn to associate and integrate modality-specific and culture-specific cues to gender in the face and voice in their own environments and generalize to specific individuals in a gender matching task. On the other hand, infants may detect any number of invariant intermodal relations that typically define gender categories (e.g., size of features, hair length, eyebrow shape, bone structure, etc.).

Infants need not integrate and associate auditory and visual information to succeed on the gender matching task. Given 8 months of perceptual learning in an environment replete with examples of gender relations, infants may become attuned to salient audiovisual invariants defining these overlapping categories. In particular, by

differentiating increasingly more specific relations over time, infants may initially detect face-voice correspondences on the basis of amodal relations (e.g., synchrony, common rhythm). This in turn may lead to abstraction of more detailed information about the nature of the synchronous face and voice (i.e., typical relations described above), and finally association of arbitrary relations such as the relation between cosmetics and jewelry with voices of higher pitch. This perceptual learning may enable more efficient abstraction of the relevant, typical relations in new instances of gender categories. By 8 months of age, infants detected audio-visual relations not by detecting temporal relations such as face-voice synchrony because these amodal invariants, which typically guide intermodal exploration, were neutralized as a basis for matching in these experiments. Rather, they must have detected these relations based on gender-specific information.

Specific aspects of audio-visual relationships may be differentially important at different times in early development. It is possible that infant's detection of certain AV relationships is limited by the function such relationships serve at particular stages in development. Speech perception is an important and computationally-demanding task for the young infant. The ability to detect audio-visual correspondences in face and voice may help infants focus on salient aspects of the speech stream and learn about the components of speech. It may thus be advantageous for infants to be unable to match gender information in the face and voice. However, once infants have developed some basic perceptual competencies with their native language, they may turn to differentiating further aspects of the face and voice such as gender.

The differentiation view is not without its critics, however. Evidence that infants respond to amodal invariants does not mean that responsiveness to intermodal relations is based entirely on such relations or that this is the only or primary way infants perceive the world (Lewkowicz, 1999). Furthermore, not all amodal information (e.g., temporal rate) is used early in development as would be predicted by the invariant detection view. The differentiation view may also place undue emphasis on information in the perceiver's world to the exclusion of possible influences that the infant's level of functional organization may have on responsiveness. Perhaps later emergence of some intersensory skills depends not on the differentiation of amodal invariants but on prior perceptual differentiation of relevant information in each modality.

In contrast to the differentiation view, the integration view (Birch & Lefford, 1963; Piaget, 1952) begins with the premise that the senses are distinct in their inputs and products and argues that infants must learn to relate the separate senses. According to Piaget, knowledge about intersensory relationships is based on the contingency of sensory outcomes on initial actions and is slowly elaborated and constructed through the infant's interaction with the world. Although this view still allows for innate sensitivity to common timing and spatial location, it is assumed that the senses operate as separate channels at birth and only with experience is input in one modality able to influence what is perceived through a different modality.

There are more recent interpretations of audio-visual perception that are consistent with an integration viewpoint. As one example, Lewkowicz (in preparation)

suggests that infants may often pay more attention to modality-specific attributes in a given modality than to the inherent amodal invariants.

> "It may be that in some cases, where the presence of an amodal invariant is highly salient ... the detection of amodal invariance occurs. In other cases, where the information in the stimulus is highly complex because of the concurrent presence of a variety of hierarchically embedded perceptual cues that are modality-specific, infants may be unable to detect the invariant relations and thus might need to learn through experience how to integrate the amodal and modality-specific information." (pp. 33-34)

The results found in this dissertation are compatible with this explanation. Amodal phonetic cues in the mouth may be highly salient compared to amodal gender cues and, thus, may faciliatate early audiovisual matching. Since visual gender involves a complex and variable mix of amodal (physiognomic), modality-specific, and culture-specific cues, a period of learning and integration may be needed.

Data from this dissertation can also be examined in light of other integration-based theories of development. According to Cohen's (1991; 1998) information-processing approach, prior to the age of 5 months, infants process only specific features of complex objects, including faces; however, between the ages of 5 and 7 months, infants start to integrate features to form meaningful wholes. The findings from this dissertation are largely consistent with this pattern. Phonetic matching could be achieved by focusing solely on an individual feature in the face (i.e., the mouth) and is seen at a relatively young age (2 months). Moreover, support for phonetic matching being based on independent features was partially obtained when the face was inverted in the phonetic matching task. Inverted faces have been shown to disrupt holistic face processing, however, infants tended to still look at the face that matched the heard

vowel. Gender matching, on the other hand, may involve integrating various different features and even culturally-specific information in the face and voice. Gender matching may also be much more variable than phonetic matching since different features are differentially reflective of gender in different individuals.

The findings from Experiment 4, when phonetic and gender information were placed in full conflict, do not fit as neatly with Cohen's (1998) predictions. Cohen assumes that if infants are presented with a task that is too difficult, the optimal strategy is to return to a lower level of processing. Thus, if infants in Experiment 4 were confused by conflicting audio-visual information, Cohen might predict that infants would fall back to matching based on vowel, a strategy that is robust at 4.5 months of age. However, the infants in Experiment 4 did not show a preference for the vowel or the gender match.

The fact that infants in this dissertation were able to match phonetic information in lips and voice at 2 months of age could also be interpreted as supporting more specific theories such as the motor theory of speech perception (Liberman & Mattingly, 1985). According to the motor theory, in order to perceive speech one must perceive a specific pattern of intended gestures which are represented in the brain as invariant motor commands that call for movement of the articulators. An innate, specialized module is proposed to lead the perceiver to this pattern of intended gestures. Central to motor theory is the notion of 'parity', the fact that articulatory structure is inherently linked to production in that particular speech sounds can only be produced by particular articulations. The importance of such a link for speech and language

necessitates an innate mechanism (i.e., specialized module). Gender, on the other hand, can be expressed in numerous different ways at different times in history and thus may need to be learned rather than based on an innate mechanism.

In summary, although the findings from this dissertation are perhaps most compatible with a differentiation view of development, they also support a number of other theories of perceptual development. This may be because the specific mechanisms underlying infant perception of intermodal equivalence are still not fully understood, but it may also be because the theories are still not precise enough to be falsifiable. It may also be the case that the questions addressed in this dissertation are not specific enough to test the theoretical positions described above. Whatever the case, I think it is important to acknowledge the possibility that the processes of integration and differentiation are both at work in intermodal development.

Recently, the co-action of both differentiation and integration processes and the interaction of factors that are both intrinsic and extrinsic to the organism has been adopted by several researchers (Gottlieb, 1991; Lewkowicz, 2000; Lickliter & Bahrick, 2000). According to this epigenetic view, reciprocal co-actions can lead to increased complexity and the elaboration of new emergent properties. Although the Gibsonian invariant detection view also considers the mutuality between the organism and its ecology to be crucial, Gottlieb's (1991) concept of co-action calls for analyses to be conducted not only at the perceptual level but at all levels of functional and structural organization.

Early-emerging abilities or sensitivities, such as phonetic matching, may be the result of a history of interaction between a genetically-initiated neural substrate and invariantly occurring species-specific experience (e.g., human speech). This epigenetic view has been adopted by infant speech perception researchers as reflecting a process of innately-guided learning (Juczyk & Bertoncini, 1988) or probabilistic epigenesis (Werker & Tees, 1999). With the human voice being a regular and reliable source of input to the fetus through both air and bone conduction (Moon et al., 1993), the neural substrate may become organized to respond preferentially to sounds that could be produced by a human vocal tract, and to process both seen and heard speech from an early age. This early sensitivity to speech would make phonetic information in the face and voice particularly salient to infants; this would inevitably lead to greater behavioural capacity and facility with speech perception which, in turn, further improves perception. The result of these reciprocal transactions is increased complexity and the elaboration of new skills and capacities. Infants' early focus on phonetic features, combined with the maturation of visual capacities, may enable them to differentiate gender-relevant information within each modality, to process more complex configural relationships in the face, and eventually match gender information with dynamic faces and voices.

Opportunities for future research in the domain of intermodal perception abound. Further inquiry into the nature of intermodal perception in newborns is greatly needed, as is investigation of the process of learning and maturation, comparative research, and how intermodal development relates to other aspects of

perceptual development such as the understanding of objects, causality, space, and time (i.e., capacities that serve as the basis for all perception).

Summary and Conclusion

In summary, my experiments with 2-, 4.5-, 6-, and 8-month-old infants demonstrate that although infants can match phonetic information in face and voice from a very young age (2 months) and even tend to do so when the face is inverted, they do not show robust evidence of gender matching with the same stimuli until 8 months of age. The ability to perceive structural correspondences between audio and visual aspects of phonetic input so early in development suggests that phonetic information may be perceived directly and may involve the detection of amodal invariants.

The results in this dissertation suggest that not all face–voice relations are equivalent. It appears that phonetic information has a special unity or salience that is not apparent in similar but non-phonetic events. Compared to phonetic matching, gender matching appears to require more experience and learning of face-voice relations that are arbitrary and culture-specific. This dissertation also provides suggestive evidence that infants may focus on individual features more than configural relationships in the face when presented with a phonetic matching task. This early attentional bias to featural/phonetic information, along with the discrimination of modality-specific gender cues, appears to set the stage for learning about culture-specific gender relations. It is not necessarily the case that young infants are "better" at

phonetic matching than at gender matching. Rather, infants appear to focus on events that are salient for them given their stage of linguistic and social development.

Appendix A

Research Methods in Infant Intermodal Perception

Methods of studying infant intermodal perception are still fairly limited. Cross-modal matching and visual habituation, and their minor variations (e.g., cross-modal transfer), will be reviewed in this section; however, other methods are also used albeit less frequently. Both the cross-modal matching and habituation techniques are based on a paradigm developed by Fantz (1961) in which changes in infants' visual attention are used to infer characteristics of cognitive processing. Visual attention is an effective dependent variable because the optical quality of the eye is very good shortly after birth and accommodative errors do not affect visual resolution (Lewis & Maurer, 1986). The human auditory system is functional at a gestational age of 25 to 27 weeks (Birnholz & Benacerraf, 1983) and infants' abilities to discriminate and categorize speech sounds become increasingly sophisticated over the first few months of life (see Werker & Tees, 1999 for a review). Research on intermodal perception has exploited the fact that looking and listening are coordinated early in life. When visual events are presented with sounds, infants' visual attention increases and they look in the direction of the sound (Muir, Humphrey, & Humphrey, 1994). More specifically, when a voice accompanies the presentation of a face, facial scanning becomes more concentrated on the mouth and eyes (Spelke & Cortelyou, 1980). It appears that the human voice, not just any interesting sound, leads to increased attention to faces.

In the *visual habituation* paradigm, a visual stimulus is presented to an infant until their visual attention decreases to a criterion level (habituation). A new visual stimulus is then presented. If the infant shows renewed visual attention, it is inferred that they can discriminate the two stimulus presentations. The cross-modal transfer method is a variation of this paradigm. Infants are first presented with a stimulus in one modality and, once habituation occurs, the familiar stimulus and a new stimulus are both presented to the infant in a different modality. If the infant looks longer at the familiar stimulus, it is assumed that the infant recognizes the stimulus across two modalities. The cross-modal transfer method is most often applied in research on visual-haptic perception. The presence of familiarity and novelty preferences makes it difficult to interpret some of the data obtained using this method (see Rose & Ruff, 1987 for a review).

Perhaps the most popular technique in research on infant audio-visual perception is the *cross-modal matching* method in which subjects are able to explore stimuli presented simultaneously in two different modalities. The most frequently used variant of this method is the Preferential Looking Technique, which was first applied to infant intermodal perception by Spelke (1976). In audio-visual matching, infants are presented with two side-by-side displays with a sound source located midway between the displays. The sound is appropriate to one of the displays; therefore, if infants look longer at the "matching" display it is inferred that they can recognize and match the auditory and visual information. After this "preference phase", a "search phase" sometimes follows in which infants are again presented with the two visual displays

along with intermittent bursts of sound appropriate to each display. Infants are expected to look longer at the sound-matched film if they have learned the relationship between the sound and the appropriate display[9].

Different procedures may tap different levels of functional organization in infants. For example, in the cross-modal transfer technique the familiarization phase provides an opportunity for learning and the time interval between stimulus presentations requires memory; therefore, this technique may reflect infants' cognitive abilities. The cross-modal matching technique is usually less demanding and may operate only on a perceptual level. Conflicting results in the intermodal perception literature may be explained in terms of experimental paradigm and level of task difficulty. A number of studies have found evidence for intermodal matching using flashing lights and tones in the visual habituation method but have failed to produce matching using the same stimuli with the preferential looking technique (e.g., Humphrey, Tees, & Werker, 1979; Humphrey & Tees, 1980; Lewkowicz, 1992). Perhaps flashing lights and tones are not interesting or meaningful enough to engage infants in a visual preference task; however, in a more demanding task, such as visual habituation, simple stimuli may be more likely to demonstrate intermodal perception. Visual preference tasks seem to be effective when stimuli are familiar and more complex.

---

[9] In different experiments, the presence of acoustic or tactile information about an object or event has influenced looking to that object in two mutually exclusive ways: (1) infants look more at the object they hear (Bahrick, 1983) or feel (Meltzoff & Borton, 1979), or (2) infants sometimes show a novelty response, looking more at the object they did not hear or touch (e.g., Gottfried, Rose, & Bridger, 1977). This second response is more commonly observed after prolonged familiarization with an object and with older infants.

Appendix B

Perceptual Abilities in 2-month-old Infants

Young infants' ability to match information in lips and voice may be limited by their visual abilities, particularly acuity, pattern recognition, and attentional focus. Newborn acuity lies within the range of 20/200 to 20/400 and does not reach near-adult levels until age 12 months (Haith, 1990). Pattern organization plays a more important role in governing visual activity after the so-called "2-month shift" than before. However, some evidence indicates that relations among pattern elements may be detected even by the newborn (Antell & Caron, 1985). Research on slightly older infants presents a mixed picture. For example, 3-month-olds behave as though they "see" a circle when presented with individual curved line segments in a circular configuration, according to Gestalt principles of good form and continuation (VanGiffen & Haith, 1984). At around the same age, infants are more likely to respond to the pattern formed by a number of elements than to the elements individually. Yet, only between 5- and 7-months of age will babies consistently behave as though they see a stationary subjective contour stimulus (Bertenthal et al., 1980).

In terms of attentional focus, some researchers have claimed that very young infants do not attend to the internal features of a visual array whereas infants older than 2 months of age do (Salapatek, 1975). Subsequent research has revealed that this may be an issue of relative saliency (Bushnell, 1979; Ganon & Swartz, 1980). Whereas external features have an advantage for the very young infant, highly salient internal

features (such as moving lips) can also be detected. Thus, infants as young as 2 months of age seem able to discriminate faces as well as detect and pay attention to moving lips.

I decided to use the preferential looking paradigm so that I could compare results between 2-month-olds and older infants and also to avoid the problem of "sticky fixation". Hood et al (1996) reported an inverted U-shape function for visual fixation measures between birth and 4 months of age. These authors found that 2-month-olds looked much longer at a grating pattern during habituation than did 1-, 3-, or 4-month-olds. This prolonged fixation behaviour has been described as "obligatory attention" (Stechler & Latz, 1966) or "sticky fixation" (Hood, 1995). Infants between 1- and 2-months of age appear to engage in long periods of staring with a fixed gaze without regularly making eye movements around the visual scene. Hood (1995) believes that this sticky looking behaviour is mainly attributable to the maturational state of the cortical and subcortical mechanisms responsible for generating orienting eye movements. The deficit may be in disengaging the eyes from a salient stimulus rather than an inability to produce an orienting eye movement. Prior to the initialization of an eye movement, inhibitory cortical input releases the superior colliculus from tonic activity, which is maintained when a visual stimulus is under gaze. Without this input, the infant has difficulty breaking fixation from a central target (Hood et al., 1996). Hood et al. (1996) found that disengagement at 2 months may be facilitated when there is a second target which can trigger orienting. Therefore, a two-choice looking paradigm such as the preferential looking technique may be a more accurate measure of looking preference than visual habituation which may exacerbate fixed attention.

References

Antell, S. & Caron, A. (1985). Neonatal perception of spatial relationships. Infant Behavior and Development, 8, 15-24.

Aslin, R.N., Saffran, J., & Newport, E.L. (1998). Computation of conditional probability statistics by 8-month-old infants. Psychological Science, 9, 321-324.

Bahrick, L.E. (1983). Infants' perception of substance and temporal synchrony in multimodal events. Infant Behavior and Development, 6, 429-451.

Bahrick, L.E. (1987). Infants' intermodal perception of two levels of temporal structure in natural events. Infant Behavior and Development, 10, 387-416.

Bahrick, L.E. (1988). Intermodal learning in infancy: Learning on the basis of two kinds of invariant relations in audible and visible events. Child Development, 59, 197-209.

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. (1998). Intermodal perception of adult and child faces and voices by infants. Child Development, 69, 1263-1275.

Bahrick, L.E. (2000). Increasing specificity in the development of intermodal perception. In D.W. Muir & A. Slater, (Eds.), Infant development: The essential readings. Oxford: Blackwell.

Bahrick, L. & Pickens, J. (1994). Amodal relations: The basis for intermodal perception and learning in infancy. In D.J. Lewkowicz & R. Lickliter (Eds.), The development of intersensory perception: Comparative perspectives (pp.205-232). Hillsdale, NJ: Earlbaum.

Banker, H. & Lickliter, R. (1993). Effects of early or delayed visual experience on perceptual development in bobwhite quail chicks. Developmental Psychobiology, 26, 155-170.

Bertenthal, B., Proffitt, D., Spetner, N. & Thomas, M. (1985). The development of infant sensitivity to biomechanical motions. Child Development, 56, 531-543.

Birch, H. & Lefford, A. (1963). Visual differentiation, intersensory integration, and voluntary motor control. Monographs of the Society for Research in Child Development, 32 (2).

Birnholz, L. & Banacerraf, R. (1983). The development of human fetal hearing. Science, 222, 516-518.

Bower, T.G.R. (1974). Development in infancy. SanFrancisco: Freeman.

Brown, E. & Perret, D. (1993). What gives a face its gender? Perception, 22, 829-840.

Bruce, V., Burton, A., Hanna, E., Healey, P., Mason, O., Coombes, A., Fright, R., & Linney, (1993). Sex discrimination: How do we tell the difference between male and female faces? Perception, 22, 131-152.

Bruce, V., Doyle, T., Dench, N., & Burton, M. (1991). Remembering facial configurations. Cognition, 38, 109-144.

Burnham, D. & Dodd, B. (1997). Auditory-visual speech perception as a direct process: The McGurk effect in infants and across languages. In D. Stork & M. Hennecke (eds.), Speechreading by Humans and Machines. Springer-Verlag.

Bushnell, I.W.R. (1979). Modification of the externality effect in young infants. Journal of Experimental Child Psychology, 28, 211-229.

Campbell, R. (1994). Audiovisual speech: Where, what, when, how? Current Psychology of Cognition, 13, 76-80.

Chronicle, E.P., Chan, M., Hawkings, C., Mason, K., Smethurst, D., Stallybrass, K.,

Westrope, K., & Wright, K. (1995). You can tell by the nose—judging sex from an isolated facial feature. Perception, 24, 969-973.

Cohen, L.B. (1991). Infant attention: An information processing approach. In M.J. Zelazo (Ed.), Newborn attention: Biological constraints and the influence of experience (pp. 1-21). Norwood, NJ: Ablex.

Cohen, L.B. (1998). An information processing approach to infant perception and cognition. In F. Simion & G. Butterworth (Eds.), The development of sensory, motor, and cognitive capacities in early infancy (pp. 277-300). East Sussex: Psychology Press.

Cohen, S. (1974). Developmental differences in infants' attentional responses to face-voice incongruity of mother and stranger. Child Development, 45, 1155-1158.

Cohen, L.B. & Cashon, C. (submitted). Do 7-month-old infants process independent features or facial configurations? Infancy.

Cooper, R. & Aslin, R.N (1989). The language environment of the young infant: Implications for early perceptual development. Canadian Journal of Psychology, 43, 247-265.

Cornell, E.H. (1974). Infants' discrimination of photographs of faces following redundant presentations. Journal of Experimental Child Psychology, 18, 98-106.

DeCasper, A.J. & Fifer, W.P. (1980). Of human bonding: Newborns prefer their mothers' voices. Science, 208, 1174-1176.

Desjardins, R.N. (1997). Audiovisual speech perception in 4-month-old infants. Unpublished doctoral dissertation. University of British Columbia, BC.

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

Echols, C., Crowhurst, M. & Childers, J. (1997). The perception of rhythmic units in speech by infants and adults. Journal of Memory and Language, 36, 202-225.

Fagan, J.F. (1972). Infants' recognition memory for faces. Journal of Experimental Child Psychology, 14, 453-476.

Fagan, J.F. (1976). Infants' recognition of invariant features of faces. Child Development, 47, 627-638.

Fagan, J.F. (1979). The origins of facial pattern recognition. In M.H. Bornstein & W. Kessen (Eds.), Psychological development from infancy: Image to intention (pp. 83-113). Hillsdale, NJ: Lawrence Erlbaum Associates, Inc.

Fantz, R.L. (1961). The origin of form perception. Scientific American, 204, 66-72.

Farah, M.J., Tanaka, J., & Drain, H.M. (1995). What causes the face inversion effect? Journal of Experimental Psychology: Human Perception and Performance, 21, 628-634.

Fowler, C. & Rosenblum, L. (1991). Perception of the phonetic gesture. In I.G. Mattingly & M. Studdert-Kennedy (Eds.), Modularity and the Motor Theory. Hillsdale, NJ: LEA.

Ghim, H. & Eimas, P.D. (1988). Global and local processing by 3- and 4-month-old infants. Perception and Psychophysics, 43, 165-171.

Ganon, H. & Swartz, E. (1980). The externality effect in neonates. Infant Behavior and Development, 6, 151-156.

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

Grant, K., Ardell, L., Kuhl, P., & Sparks, D. (1986). The transmission of prosodic information via an electrotactile speechreading aid. Ear and Hearing, 7, 328-335.

Green, K. (1994). The influence of an inverted face on the McGurk effect. Journal of the Acoustical Society of America, 95, 3014.

Green, K. & Kuhl, P.K. (1991). Integral processing of visual place and auditory voicing information during phonetic perception. Journal of Experimental Psychology: Human

Perception and Performance, 17, 278-288.
Gottfried, A., Rose, S. & Bridger, W. (1977). Cross-modal transfer in human infants. Child Development, 48, 118-123.
Gottlieb, G. (1991).
Haith, M. (1990). Progress in the understanding of sensory and perceptual processes in early infancy. Merrill-Palmer Quarterly, 36, 1-26.
Hirsch-Pasek, K. & Golinkoff, R.M. (1992). Skeletal supports for grammatical learning: what infants bring to the language learning task. In L.P. Lipcott & C. Rovee-Collier (Eds.), Advances in infancy research (Vol. 8., pp. 299-338). Norwood, NJ: Ablex.
Hood, B. (1995). Disengaging visual attention in the infant and adult. Infant Behavior and Development, 16, 405-422.
Hood, B., Murray, L., King, F., Hooper, R. (1996). Habituation changes in early infancy: Longitudinal measures from birth to 6 months. Journal of Reproductive and Infant Psychology, 14, 177-185.
Humphrey, K. & Tees, R.C. (1980). Auditory-visual coordination in infancy: Some limitations of the preference methodology. Bulletin of the Psychonomic Society, 16, 213-216.
Humphrey, K., Tees, R.C., & Werker, J. (1979). Auditory-visual integration of temporal relations in infants. Canadian Journal of Psychology, 33, 347-352.
Hunter, M. & Ames, E. (1988). A multi-factor model of infant preferences for novel and familiar stimuli. In L.P. Lipsitt & C. Rovee-Collier (Eds.), Advances in infancy research (Vol.5, pp.69-95). Norwood, NJ: Ablex.
Jordan, T.R. & Bevan, K. (1997). Seeing and hearing rotated faces: Influences of facial orientation on visual and audiovisual speech recognition. Journal of Experimental Psychology: Human Perception and Performance, 23, 388-403.
Jusczyk, P.M. & Aslin, R. (1995). Infants' detection of the sound patterns of words in fluent speech. Cognitive Psychology, 29, 1-23.
Jusczyk, P. & Bertoncini, J. (1988). Viewing the development of speech perception as an innately-guided learning process. Language and Speech, 31, 217-238.
Kellman, P.J. & Arterberry, M.E. (1998). The cradle of knowledge: Development of perception in infancy, Cambridge, MA: MIT Press.
Kestenbaum, R. & Nelson, C.A. (1990). The recognition and categorization of upright and inverted emotional expressions by 7-month-old infants. Infant Behavior and Development, 13, 497-511.
Kuhl, P.K. (1979). Speech perception in early infancy: Perceptual constancy for spectrally dissimilar vowel categories. Journal of the Acoustical Society of America, 66, 1668-1679.
Kuhl, P.K. (1983). Perception of auditory equivalence classes for speech in early infancy. Infant Behaviour and Development, 6, 263-285.
Kuhl, P.K. (1991). Human adults and human infants show a perceptual magnet effect for the prototypes of speech categories, monkeys do not. Perception & Psychophysics, 50, 93-107.
Kuhl, P.K. & Meltzoff, A.N. (1982). The bimodal development of speech in infancy. Science, 218, 1138-1141.
Kuhl, P.K. & Meltzoff, A.N. (1984). The bimodal representation of speech in infants. Infant Behavior and Development, 7, 361-381.
Kuhl, P.K. & Meltzoff, A.N. (1988). Speech as an intermodal object of perception. In A. Yonas (Ed.), Perceptual development in infancy: The Minnesota Symposia on Child

Psychology (Vol. 20, pp. 235-266). Hillsdale, NJ: Earlbaum.
Kuhl, P.K. & Miller, J. (1982). Discrimination of auditory target dimensions in the presence or absence of variation in a second dimension by infants. Perception & Psychophysics, 31, 279-292.
Kuhl, P.K., Williams, K.A., & Meltzoff, A.N. (1991). Cross-modal speech perception in adults and infants using nonspeech auditory stimuli. Journal of Experimental Psychology: Human Perception and Performance, 17, 829-840.
Ladefoged, P. (1993). A course in phonetics. 3rd Edition. Forth Worth, NY: Harcourt-Brace.
Laskey, R.E., Klein, R.E., & Martinez, S. (1974). Age and sex discrimination in five- and six-month-old infants. Journal of Psychology, 88, 317-324.
Legerstee, M. (1990). Infants use multimodal information to imitate speech sounds. Infant Behavior and Development, 13, 343-354.
Leinbach. M.D. & Fagot, B. (1993). Categorical habituation to male and female faces: Gender schematic processing in infancy. Infant Behavior and Development, 16, 317-332.
Lewis, T. & Maurer, D. (1986). Preferential looking as a measure of visual resolution in infants and toddlers. Child Development, 57, 1062-1075.
Lewkowicz, D. (1985). Developmental changes in infants' visual response to temporal frequency. Developmental Psychology, 21, 858-865.
Lewkowicz, D. (1992). Infants' responsiveness to the auditory and visual attributes of a sounding/moving stimulus. Perception & Psychophysics, 52, 519-528.
Lewkowicz, D. (1996). 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. (1998). Infants' response to the audible and visible properties of the human face: II. Discrimination of differences between singing and adult-directed speech. Developmental Psychology, 32, 261-274.
Lewkowicz, D. (August, 1999). Infants' perception of the audible, visible, and bimodal attributes of talking and singing faces. Proceedings of the Audio-Visual Speech Processing Conference, University of California, Santa Cruz.
Lewkowicz, D. (2000). The development of intersensory temporal perception: An epigenetic systems/limitations view. Psychological Bullentin, 126, 281-308.
Lewkowicz, D. (in prep). Heterogeneity and heterochrony in the development of intersensory perception.
Lewkowicz, D.J. & Lickliter, R. (1994). Insights into mechanisms of intersensory development. In D.J. Lewkowicz & R. Lickliter (Eds.), The development of intersensory perception: Comparative perspectives (pp. 403-413). Hillsdale, NJ: Earlbaum.
Lewkowicz, D.J. & Turkewitz, G. (1980). Cross-modal equivalence in early infancy: Auditory-visual intensity matching. Developmental Psychology, 16, 597-607.
Liberman, A.M. & Mattingly, I.G. (1985). Motor theory of speech perception revised. Cognition, 21, 1-36.
Lickliter, R. & Bahrick, L.E. (2000). The development of infant intersensory perception: Advantages of a comparative convergent-operations approach. Psychological Bulletin, 126, 260-280.
Lyons-Ruth, K. (1977). Bimodal perception in infancy: Responses to auditory-visual incongruity. Child Development, 48, 820-827.
MacKain, K., Studdert-Kennedy, M., Spieker, S., & Stern, D. (1983). Infant intermodal speech perception is a left hemisphere function. Science, 219, 1347-1349.

Marean, G.C., Kuhl, P., & Werner, L.A. (1992). Vowel categorization by very young infants. Developmental Psychology, 28, 396-405.

Massaro, D.W. & Cohen, M. (1996). Perceiving speech from inverted faces. Perception and Psychophysics, 58, 1047-1065.

McClure, E. (2000). A meta-analytic review of sex differences in facial expression processing and their development in infants, children, and adolescents. Psychological Bulletin, 126, 424-453.

McGurk, H. & MacDonald, J.W. (1976). Hearing lips and seeing voices. Nature, 264, 746-748.

Meltzoff, A.N. & Borton, C. (1979). Intermodal matching by human neonates. Nature, 282, 403-404.

Meltzoff, A.N. & Moore, K. (1983). Newborn infants imitate adult facial gestures. Child Development, 54, 702-709.

Meltzoff, A.N. & Kuhl, P.K. (1994). Faces and speech: Intermodal processing of biologically relevant signals in infants and adults. In D.J. Lewkowicz & R. Lickliter (Eds.), The development of intersensory perception: Comparative perspectives (pp.39-55). Hillsdale, NJ: Earlbaum.

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

Miller, C.L. & Horowitz, F.D. (1980, April). Integration of auditory and visual cues in speaker classification by infants. Paper presented at the International Conference on Infant Studies, New Haven, CT.

Moon, C., Cooper, R., & Fifer, W. (1993). Infants prefer native language. Infant Behavior and Development, 16, 495-500.

Morrongiello, B. (1994). Effects of colocation on auditory-visual interactions and cross-modal perception in infants. In D.J. Lewkowicz & R. Lickliter (Eds.), The development of intersensory perception: Comparative perspectives (pp.39-55). Hillsdale, NJ: Earlbaum.

Morrongiello, B., Fenwick, K., & Nutley, G. (1998). Crossmodal learning in newborn infants: Inferences about properties of auditory-visual events. Infant Behavior and Development, 21, 543-554.

Muir, D. W., Humphrey, D. & Humphrey, G.K. (1994). Pattern and space perception in young infants. Spatial Vision, 8, 141-165.

Nelson, C. & Ludemann, P. (1989). Past, current, and future trends in infant face perception research. Canadian Journal of Psychology, 43, 183-198.

Oller, D.K. (1986). The emergence of speech sounds in infancy. In G. Yeni-Komshian, J. Kavanagh, C. Ferguson (Eds.), Child Phonology (pp. 92-112). NewYork: Academic.

Patterson, M. & Werker, J.F. (1999). Matching phonetic information in lips and voice is robust in 4.5-month-old infants. Infant Behavior and Development, 22, 237-247.

Peterson, G.E. & Barney, H.L. (1952). Control methods used in a study of the vowels. Journal of the Acoustical Society of America, 24, 175-184.

Piaget, J. (1952). The origins of intelligence in children. NewYork: International Universities Press.

Polka, L. & Werker, J.F. (1994). Developmental changes in perception of nonnative vowel contrasts. Journal of Experimental Psychology: Human Perception and Performance, 20, 421-435.

Poulin-Dubois, D., Serbin, L., Kenyon, B., & Derbyshire, A. (1994). Infants' intermodal knowledge about gender. Developmental Psychology, 30, 436-442.

Rock, I. (1974). The perception of disoriented figures. Scientific American, 230, 78-85.

Rose, S.A. & Ruff, H.A. (1987). Cross-modal abilities in human infants. In J.D. Osofsky (Ed.),

Handbook of infant development. (pp. 318-362). NY: Wiley.
Rosenblum, L., Smuckler, M., & Johnson, J. (1997). The McGurk effect in infants. Perception and Psychophysics, 59, 347-357.
Rosenblum, L., Yakel, D., & Green, K. (2000). Face and mouth inversion effects on visual and audiovisual speech perception. Journal of Experimental Psychology: Human Perception and Performance, 26, 806-819.
Saffran, J. (April, 2001). Statistical learning of auditory and linguistic patterns. Paper presented at the Society for Research on Child Development. Minneapolis, MN.
Salapatek, P. (1975). Saccadic localization of visual targets by the very young infant. Perception & Psychophysics, 17, 293-302.
Schwarzer, G. (2000). Development of face processing: The effect of face inversion. Child Development, 71, 391-401.
Searcy, J. & Bartlett, J. (1996). Inversion and categorization of component and spatial-relational information in faces. Journal of Experimental Psychology: Human Perception and Performance, 22, 904-915.
Sergent, J. (1984). An investigation into component and configural processes underlying face perception. British Journal of Psychology, 75, 221-242.
Smith, L.B. (1989). A model of perceptual categorization in children and adults. Psychological Review, 96, 125-144.
Spear, N. & McKinzie, D. (1994). Intersensory integration in the infant rat. In D.J. Lewkowicz & R. Lickliter (Eds.), The development of intersensory perception: Comparative perspectives (pp. 133-161). Hillsdale, NJ: Earlbaum.
Spelke, E. (1976). Infants' intermodal perception of events. Cognitive Psychology, 8, 553-560.
Spelke, E. & Cortelyou, A. (1980). Perceptual aspects of social knowing: Looking and listening in infancy. In M. Lamb & L. Sherrod (Eds.), Infant Social Cognition. NY: Freeman.
Spelke, E. & Owsley, C.J. (1979). Intermodal exploration and knowledge in infancy. Infant Behavior and Development, 2, 13-27.
Stechler, G. & Latz, E. (1966). Some observations on attention and arousal in the human infant. Journal of American Academy of Child Psychiatry, 5, 517-525.
Symons, L. & Tees, R.C. (1990). An examination of the intramodal and intermodal behavioral consequences of long-term vibrassae removal in rats. Developmental Psychobiology, 23, 849-867.
Thelen, E. & Smith, L. (1994). A dynamic systems approach to the development of cognition and action. Cambridge, MA: MIT Press.
Tincoff, R. & Jusczyk, P. (1999). Some beginnings of word comprehension in 6-month-olds. Psychological Science, 10, 172-175.
Trehub, S. (1973). Infants' sensitivity to vowel and tonal contrasts. Developmental Psychology, 9, 91-96.
Vatikiotis-Bateson, (2001, March). Multimodal speech, redundancy, and communication. Paper presented at The Listening Brain conference. University of British Columbia, Vancouver, BC.
Valentine, T. & Bruce, V. (1985). What's up? The Margaret Thatcher illusion revisited. Perception, 14, 515-516.
VanGiffen, K. & Haith, M. (1984). Infant visual response to Gestalt geometric forms. Infant Behavior and Development, 7, 335-346.
Walker, A.S. (1982). Intermodal perception of expressive behaviors by human infants. Journal of Experimental Child Psychology, 33, 514-535.
Walker-Andrews, A.S. (1997). Infants' perception of expressive behaviors: Differentiation of

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

Walker-Andrews, A. (1994). Taxonomy for intermodal relations. In D.J. Lewkowicz & R. Lickliter (Eds.), The development of intersensory perception: Comparative perspectives (pp.39-55). Hillsdale, NJ: Earlbaum.

Walker-Andrews, A.S., Bahrick, L.E., Raglioni, S.S., & Diaz, I. (1991). Infants' bimodal perception of gender. Ecological Psychology, 3, 55-75.

Walton, G.E. & Bower, T.G.R. (1993). Amodal representation of speech in infants. Infant Behavior and Development, 16, 233-243.

Werker, J.F. & Tees, R.C. (1992). The organization and reorganization of human speech perception. Annual Review of Neuroscience, 15, 377-402.

Werker, J.F. & Tees, R.C. (1999). Influences on infant speech processing: Toward a new synthesis. Annual Review of Psychology, 50, 509-535.

Yin, R.K. (1969). Looking at upside-down faces. Journal of Experimental Psychology: Human Learning and Memory, 7, 181-190.

Younger, B.A. & Fearing, D. (1999). Parsing items into separate categories: Developmental change in infant cognition. Child Development, 70, 291-303.

Figure Captions

Figure 1. Visual stimuli used in Experiments 1 – 9 (In Experiment 7 the entire face was inverted and in Experiment 8 only the mouth was inverted).

Figure 2. Examples of experimental design for Experiments 1-9.

Figure 3. Mean percentage of total looking time (PTLT) to the face that matched the gender of the heard sound as a function of infants' age in months. Error bars represent standard error.

Figure 4. Mean percentage of total looking time to the face that matched the vowel and/or gender of the heard voice for the conflict studies (Experiments 2-4). Error bars represent standard error.

Figure 5. Mean percentage of total looking time to the face that matched the heard vowel at 2 months (Experiment 9) and at 4.5 months (Patterson & Werker, 1999) of age. Error bars represent standard error.

Figure 1. Examples of visual stimuli.

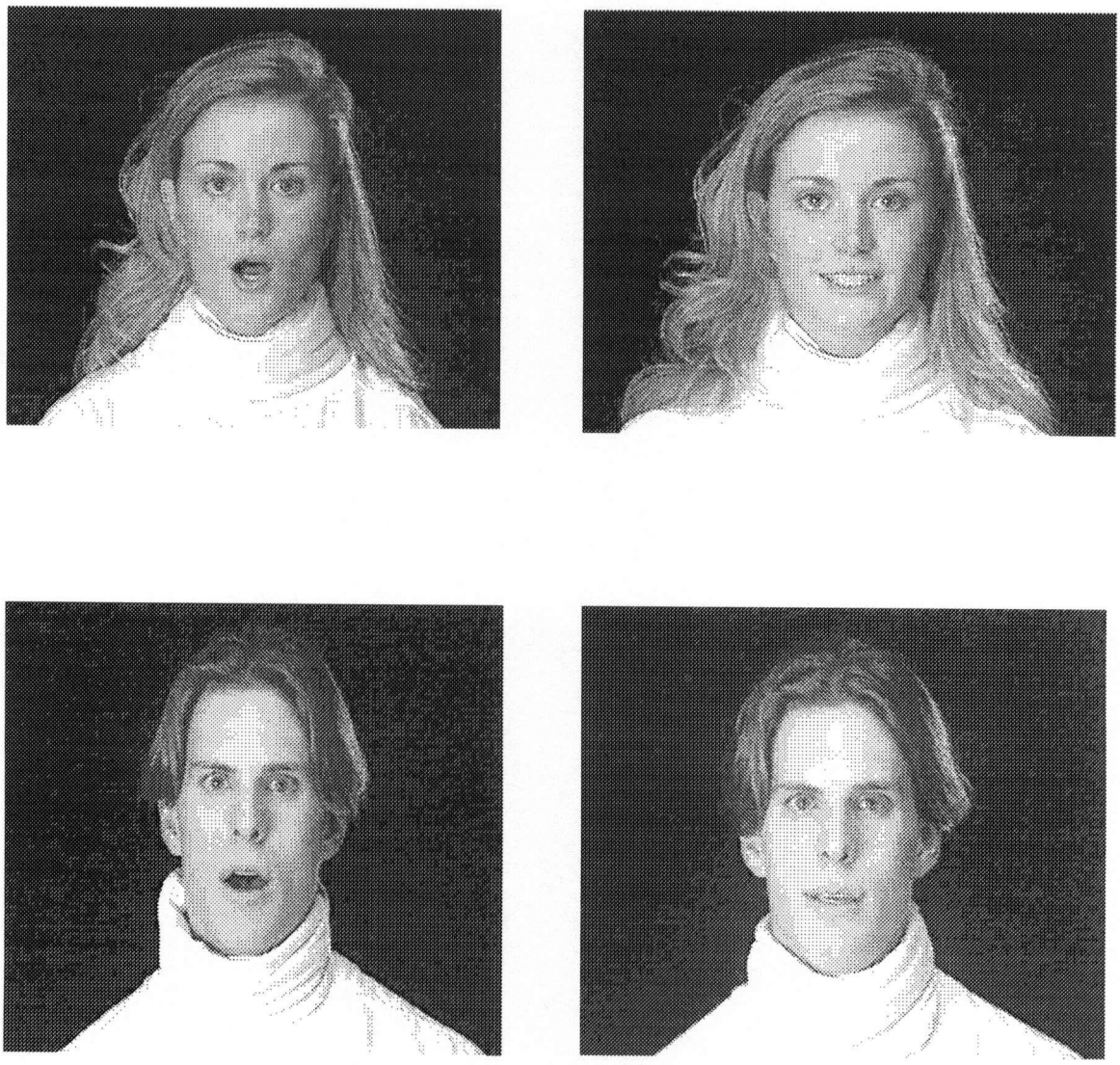
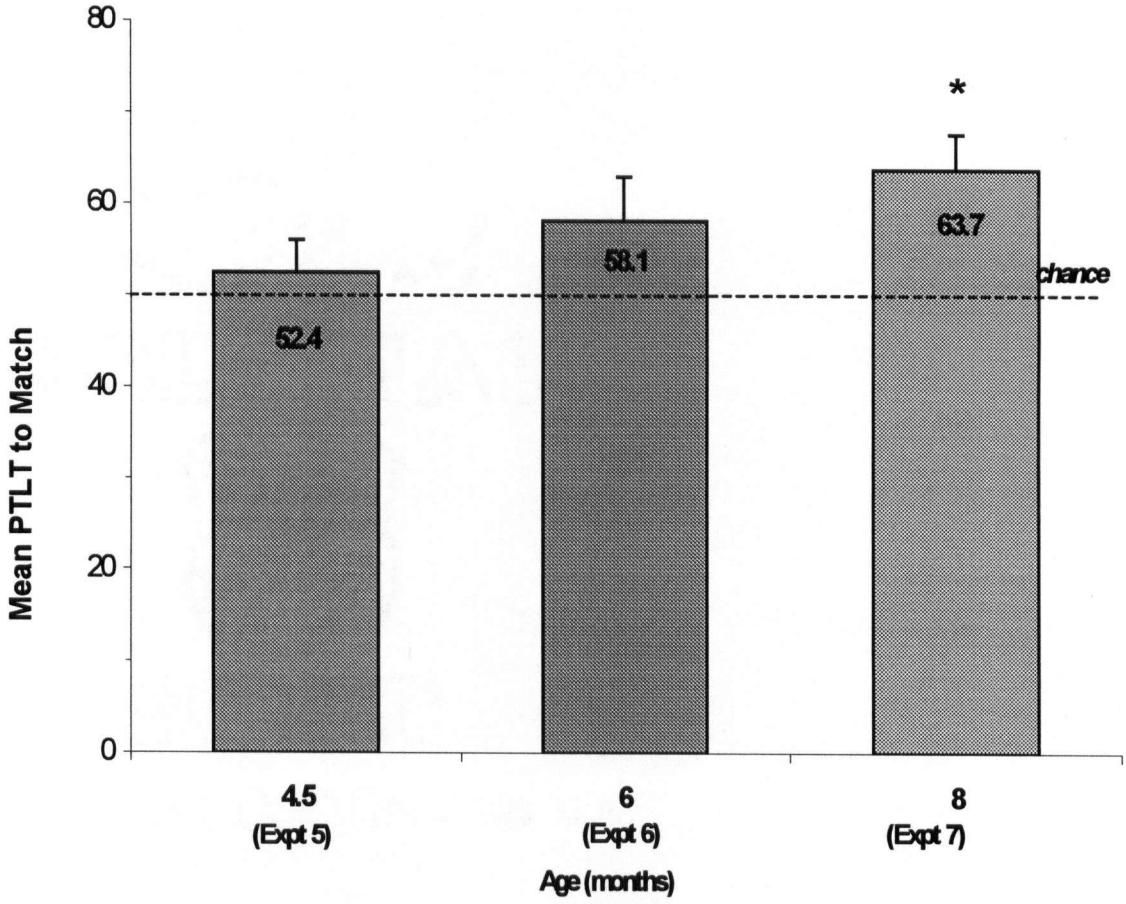


Figure 2. Examples of experimental design.

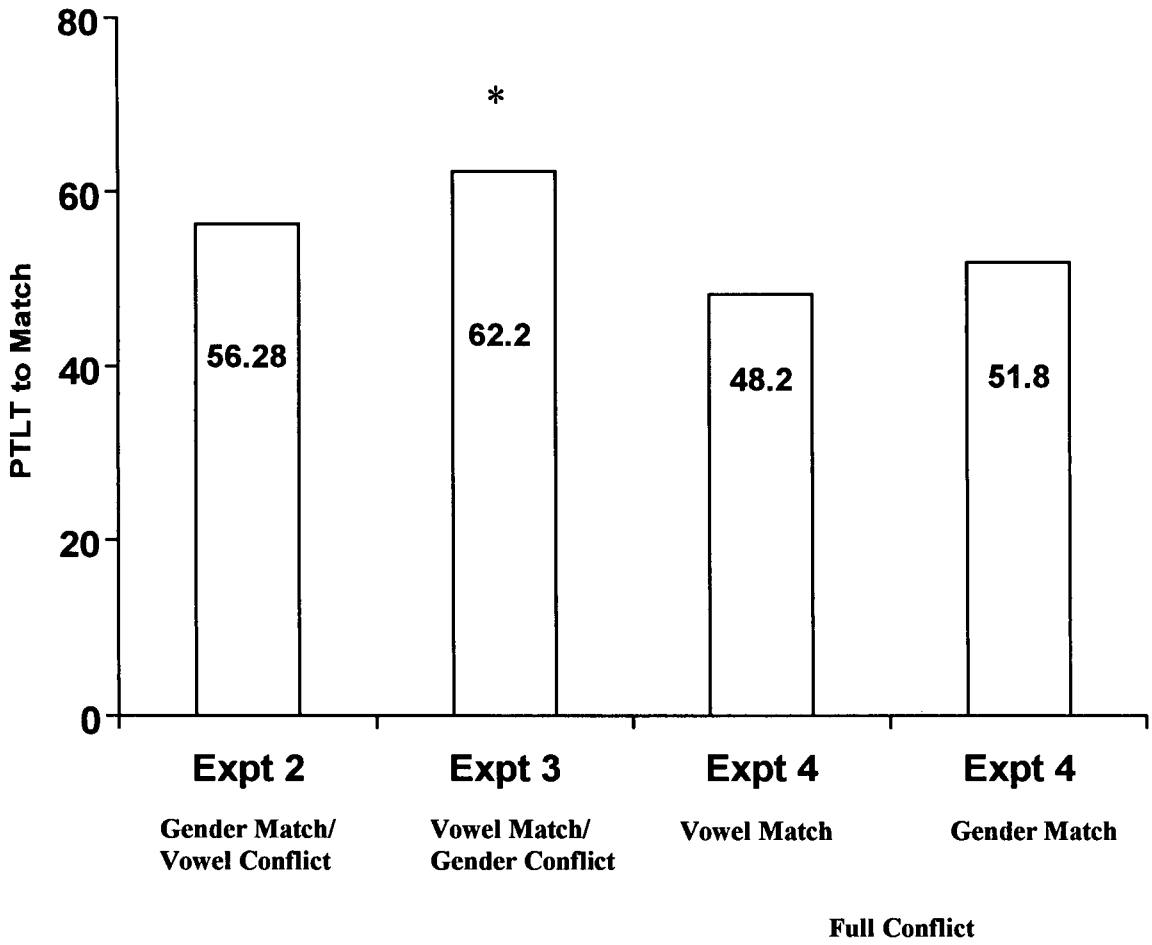
	FAMILIARIZATION			TEST
TIME (sec)	**0**	**4**	**7**	**10-129**
EXPERIMENTS 1, 5, 6: Gender matching				
LEFT TV	F/a/	---	---	F/a/
RIGHT TV	---	M/a/	---	M/a/
SOUND	---	---	---	F/a/
EXPERIMENT 2: Gender matching with discrepant vowel				
LEFT TV	F/a/	---	---	F/a/
RIGHT TV	---	M/a/	---	M/a/
SOUND	---	---	---	M/i/
EXPERIMENT 3: Vowel matching with discrepant gender				
LEFT TV	F/a/	---	---	F/a/
RIGHT TV	---	F/i/	---	F/i/
SOUND	---	---	---	M/a/
EXPERIMENT 4: Vowel and gender conflict				
LEFT TV	F/a/	---	---	F/a/
RIGHT TV	---	M/i/	---	M/i/
SOUND	---	---	---	F/i/
EXPERIMENTS 7-9: Phonetic matching				
LEFT TV	F/a/	---	---	F/a/
RIGHT TV	---	F/i/	---	F/i/
SOUND	---	---	---	F/a/

Figure 3. Mean percentage of total looking time (PTLT) to the face that matched the gender of the heard sound as a function of infants' age in months.



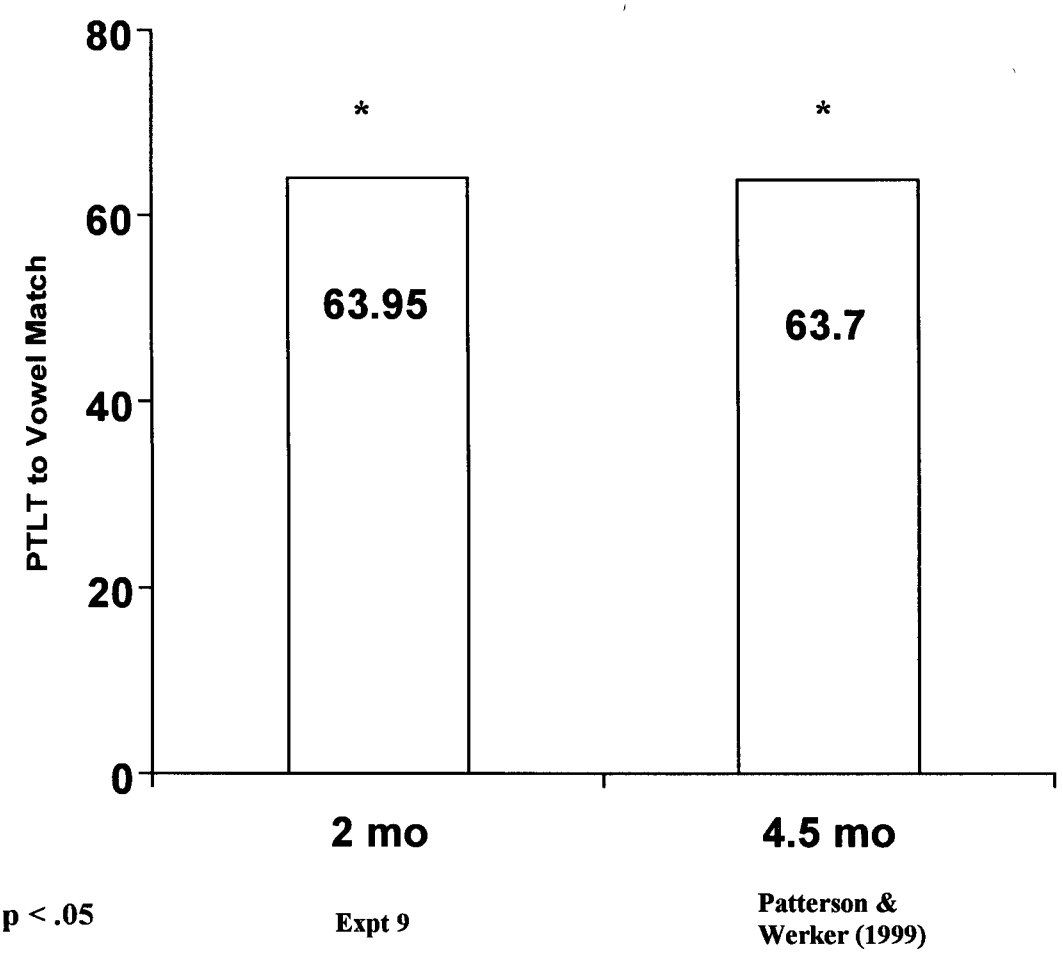
* p<.05

Figure 4. Mean percentage of total looking time to the face that matched the vowel and/or gender of the heard voice for the conflict studies (Experiments 2-4).



* p > .05

Figure 5. Mean percentage of total looking time to the face that matched the heard vowel at 2 months (Experiment 9) and at 4.5 months (Patterson & Werker, 1999) of age.



Table

Corrected proportions of looking time in Experiments 5 and 6 based on side of match

Age (months)	Left Side	Right Side	Total corrected looking to match
6	.61	.56	.59
8	.67	.64	.65*

* $p<.05$, one-tailed test

Note. Due to significant side bias in Experiments 5 and 6, Humphrey and Tees' (1980) correction procedure was applied to the data. In both experiments, the correction procedure did not change the overall results.